E77-10062

NASA CR-
ERIM 109600-67-F

NASA CR-
151002

NASA

Final Report

SYSTEM FOR ANALYSIS OF LANDSAT AGRICULTURAL DATA:

Automatic Computer-Assisted Proportion Estimation of Local Areas

R.J. KAUTH AND G.S. THOMAS
Infrared and Optics Division

MAY 1976

(E77-10062) SYSTEM FOR ANALYSIS OF LANDSAT AGRICULTURAL DATA: AUTOMATIC COMPUTER-ASSISTED PROPORTION ESTIMATION OF LOCAL AREAS Final Report, 15 May 1975 - 14 May 1976 (Environmental Research Inst. of G3/43

N77-14561
HC A06
MF A01
Unclas
00062

Original photography may be purchased from:
EROS Data Center
10th and Dakota Avenue
Sioux Falls, SD 57198

Prepared for

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Johnson Space Center
Earth Observations Division
Houston, Texas 77058
Contract No. NAS9-14123, Task 12
Technical Monitor: Dr. A. Potter/TF3

ENVIRONMENTAL
RESEARCH INSTITUTE OF MICHIGAN
FORMERLY WILLOW RUN LABORATORIES, THE UNIVERSITY OF MICHIGAN
BOX 618•ANN ARBOR•MICHIGAN 48107

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No. NASA CR-ERIM 109600-67-F	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle System for Analysis of Landsat Agricultural Data: Automatic Computer Assisted Proportion Estimation of Local Areas		5. Report Date May 1976 6. Performing Organization Code
7. Author(s) R. J. Kauth and G. S. Thomas		8. Performing Organization Report No. 109600-67-F
9. Performing Organization Name and Address Environmental Research Institute of Michigan Infrared & Optics Division P. O. Box 618 Ann Arbor, Michigan 48107		10. Work Unit No. Task 12 11. Contract or Grant No. NAS9-14123 13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address National Aeronautics & Space Administration Johnson Space Center Houston, Texas 77058		Final Technical Report May 15, 1975 through May 14, 1976 14. Sponsoring Agency Code

15. Supplementary Notes

The work was performed for the Earth Observations Division.
Dr. Andrew Potter (TF3) was the technical monitor.

16. Abstract

The purpose of this work is to develop computer techniques for assisting an Analyst-Interpreter (AI) in the task of training field identification. The result of the work has been to develop an integrated man-machine approach to the problem of local proportion estimation in large scale agricultural remote sensing systems, of which LACIE (Large Area Crop Inventory Experiment) is an example. The approach builds on the current LACIE system structure.

Local proportion estimation has two major aspects; the organization of the data of the sample segment in such a way as to simplify the task of AI identification of the training fields and to integrate the AI designations more closely into the proportion estimation process; and the actual process of designating training fields whether by AI, by computer, or by a joint effort of both. A partial system for performing the second task is described, and the conceptual basis of the approach is explored. A complete system for performing the first of the above tasks is described; the system has been implemented using ground truth in lieu of AI designations; and a few examples of proportion estimates are given.

17. Key Words		18. Distribution Statement
Remote Sensing Pattern Recognition Data Structure Landsat Agricultural Surveys Crop Identification Crop Phenology	Image Interpretation Spectral-Spatial-Temporal Clustering Soil Brightness Effects Atmosphere Effects Man-Machine Integration	Initial distribution is listed at the end of this document.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	viii + 92	

PREFACE

This report describes part of a comprehensive and continuing program of research concerned with advancing the state-of-the-art in remote sensing of the environment from aircraft and satellites. The research is being carried out for NASA's Lyndon B. Johnson Space Center, Houston, Texas, by the Environmental Research Institute of Michigan (ERIM). The basic objective of this multidisciplinary program is to develop remote sensing as a practical tool to provide the planner and decision-maker with extensive information quickly and economically.

Timely information obtained by remote sensing can be important to such people as the farmer, the city planner, the conservationist, and others concerned with problems such as crop yield and disease, urban land studies and development, water pollution, and forest management. The scope of our program includes:

1. extending the understanding of basic processes.
2. discovering new applications, developing advanced remote-sensing systems, and improving automatic data processing to extract information in a useful form.
3. assisting in data collection, processing, analysis, and ground-truth verification.

The research described herein was performed under NASA Contract NAS9-14123, Task 12, and covers the period from May 15, 1975 through May 14, 1976. Andrew Potter (TF3) was the NASA Contract Technical Monitor. The program was directed by Richard R. Legault, Vice-President of ERIM and Head of the Infrared and Optics Division, Jon D. Erickson, Head of the Information & Analysis Department, and Richard F. Nalepka, Principal Investigator and Head of the Multispectral Analysis Section.

The authors acknowledge the excellent programming support of A. Pentland. The authors recognize and appreciate the illumination provided by H. Horwitz and J. Colwell of ERIM, and M. Trichel and R. Heydorn of Johnson Space Center. In addition, numerous other individuals contributed to the general concepts of this report.

CONTENTS

FIGURES

FIGURES

TABLES

1
SUMMARY

The purpose of this work is to develop computer techniques for assisting an Analyst-Interpreter (AI) in the task of training field identification. The result of the work has been to develop an integrated man-machine approach to the problem of local proportion estimation in large scale agricultural remote sensing systems, of which LACIE (Large Area Crop Inventory Experiment) is an example. The approach builds on the current LACIE system structure.

Local proportion estimation has two major aspects; the organization of the data of the sample segment in such a way as to simplify the task of AI identification of the training fields and to integrate the AI designations more closely into the proportion estimation process; and the actual process of designating training fields whether by AI, by computer, or by a joint effort of both. A partial system for performing the second task is described, and the conceptual basis of the approach is explored. A complete system for performing the first of the above tasks is described; the system has been implemented using ground truth in lieu of AI designations; and a few examples of proportion estimates are given.

2
INTRODUCTION

The Classification and Mensuration Subsystem of LACIE covers three principle functional areas.

1. Training Field Identification
2. Training Statistics
3. Proportion Estimation.

The training field identification function in LACIE is carried out by Analyst-Interpreters (AI's). Training statistics may be extracted from the identified training fields and be used locally (in the same sample segment) for proportion estimation in the LACIE II processor; or training statistics may be extracted from the identified training fields, be modified in some way, and be used non-locally (in some other sample segment) for proportion estimation.

The work accomplished under this task was directed to the objective of assisting the AI's in performance of the training field identification function, either by helping AI's identify wheat training fields, or by direct computer identification of wheat training fields. However, a slight generalization leads to the statement that the objective is to create an AI-computer system for performing local proportion estimation.

The general approach chosen for this task was to examine the functions which must be performed to identify wheat fields at a level of generality that would be valid whether either a man or a computing machine were carrying out the task; to identify the functions which might best be carried out by the computing machine in the near or intermediate term future; and to attempt to implement some or all of these functions.

The functions which were considered were restricted to those which could be carried out using data which could reasonably be obtained in an

operational large area crop inventory system utilizing Landsat data, namely: the Landsat image data; historical crop calendars, cropping practices, crop acreages, meteorological data; current crop calendars, meteorological station and meteorological satellite data; and various spectral and spatial features drawn from the Landsat images. The Landsat images as used by AI's or as processed for proportion estimation are termed image data; all other types of data will be termed ancillary data.

Since there are choices available in precisely what types and sources of data to use, there is no unique set of functions which will be necessary and sufficient to identify wheat fields or to perform an estimate of wheat proportion. Our approach will be to identify as many functions as we can and place them in a reasonable order in a function flow diagram.

The observation of the behavior of the AI's within the context of the LACIE system is a fruitful source for identifying specific functions. Other portions of the LACIE system can provide ideas as well. In addition, we want to be alert to the possible existence of useful functions not being carried on anywhere in the LACIE system.

3
FUNCTION ANALYSIS

Within the LACIE system, or a follow-on agricultural remote sensing system, we have thought of the AI as providing starting points for both local and non-local recognition. However, there has been no notable success in the non-local recognition even though rather significant success has been attained in local recognition (when particular useful combinations of passes were available). At this point in time, we perceive two broad reasons for failure of non-local recognition. One of these is the influence of external effects such as viewing angle, atmosphere haze, and water vapor. It has been well demonstrated now that failing to correct these effects will result in essentially random proportion estimations [14]. These effects can be corrected by a variety of means being pursued throughout the SR&T effort under the title of signature extension.

The second broad reason for failure of non-local recognition can be seen in terms of a sampling problem. The individual fields or regions within a sample segment can be modeled as being the result of some random draw from a large region called a partition. Within this large region the variability of signatures may be large, but the statistics of the variability are stationary. The collection of fields actually occurring in any particular sample segment is thus a small sample from a broad distribution. The chances are small that the sample statistics of the training fields picked by the AI in the training (local) sample segment will match the sample statistics of the fields in the recognition segment.

One might ask why it is that local recognition seems to work rather well. Surely the same sampling differences occurs within sample segments as between them; why is it not equally true that the chances are small that the sample statistics of the training fields picked by the AI will match the sample statistics of the remainder of the fields in the local

sample segment? The reason is that the AI deliberately tries to pick fields which are representative of the local segment. In effect, he analyzes the statistics of the entire segment and then picks individual fields which are representative of the entire segment. In order to obtain the same success in non-local recognition, the AI would have to see the non-local segment and pick fields from the training segment which were representative of the non-local segment. In most cases such fields would by chance not exist. Even if they did, the AI could not examine every recognition segment.

Evidently, the key is to select fields which are representative of the entire partition. In order to do this, it will be necessary to find training fields in many sample segments and to correct external effects among and between these segments in order to establish a composite training data base for the partition.

The above paragraphs have been directed to the subject of non-local recognition or signature extension. Let us return now to the task of local identification of crops.

If the AI must be involved in every local segment, we at least would like to be able to relieve him of some of the burden through computer techniques. We see the goal of a gradual reduction in the AI work load at each site until finally he is in the role of monitoring and checking computer results. We are encouraged that this goal may be a practical one by the early results of the delta classifier [12].

With the above paragraphs as an introduction, we now briefly review the functions which the AI must perform in order to identify field crops, and indicate the corresponding computer support functions which we have developed.

Table 1 lists the AI functions which we perceive at this time. The associated computer support functions are also shown. Notice that

TABLE 1. LIST OF CONCEPTUALIZED AI FUNCTIONS AND RELATED COMPUTER SUPPORT FUNCTIONS

AI FUNCTIONS	ASSOCIATED COMPUTER SUPPORT FUNCTIONS
Locate fields. Fix fields.	Isolate fields and flag all pixels belonging to the particular fields. Present to AI.
Estimate field characteristics.	Compute spectral-temporal statistics of each field.
Tentatively identify fields.	Suggest field identification to AI for his verification.
Review field characteristics and select a representative set of fields.	Cluster fields into higher groups. Select a representative set of fields.
Final field identification.	Suggest field identification to AI for his spot checking.

one computer support function, isolate and flag field pixels, is shown opposite two separate AI functions. In general, there is no exact correspondence between the beginning and end of AI steps and computer steps.

It is not necessary to implement all computer support functions in order to benefit from some of them. However, each different partial implementation requires a different interface with the rest of CAMS. A progression of examples is shown in Figures 1(a) through 1(e). We show this entire set of figures to demonstrate the gradual changes in system function which are possible without premature commitment to the final stage.

Figure 1(a) shows the current configuration with emphasis on the AI. The ancillary data in the form of crop calendars and cropping practices comes to the AI along with film products from the Landsat imagery.

In Figure 1(b) an additional data product is delivered to the AI, namely a transparency in which the flagged field center pixels are printed and the non-flagged boundary pixels are suppressed. Using a lined overlay, the AI can designate the center of any group of pixels by approximate line and point. The computer then matches up the line and point with the mean line and point of a field. These actions replace the drawing of fields and the extraction of vertices. The flagging of field pixels can be accomplished based on all of the passes available to that date, i.e., field isolation can be based on multitemporal data.

In order to define fields, the computer will be calculating field statistics anyway, so these statistics can be saved and used for training the LACIE processor, rather than recalculating statistics from the training fields. This option is shown in Figure 1(c).

In Figure 1(d), the AI is provided with an additional source of information, namely a list of groupings of fields according to their multitemporal/multispectral characteristics. This list supplements the AI's own visual characterization of the entire scene and assists him in choosing representative fields. For example, after choosing a set of

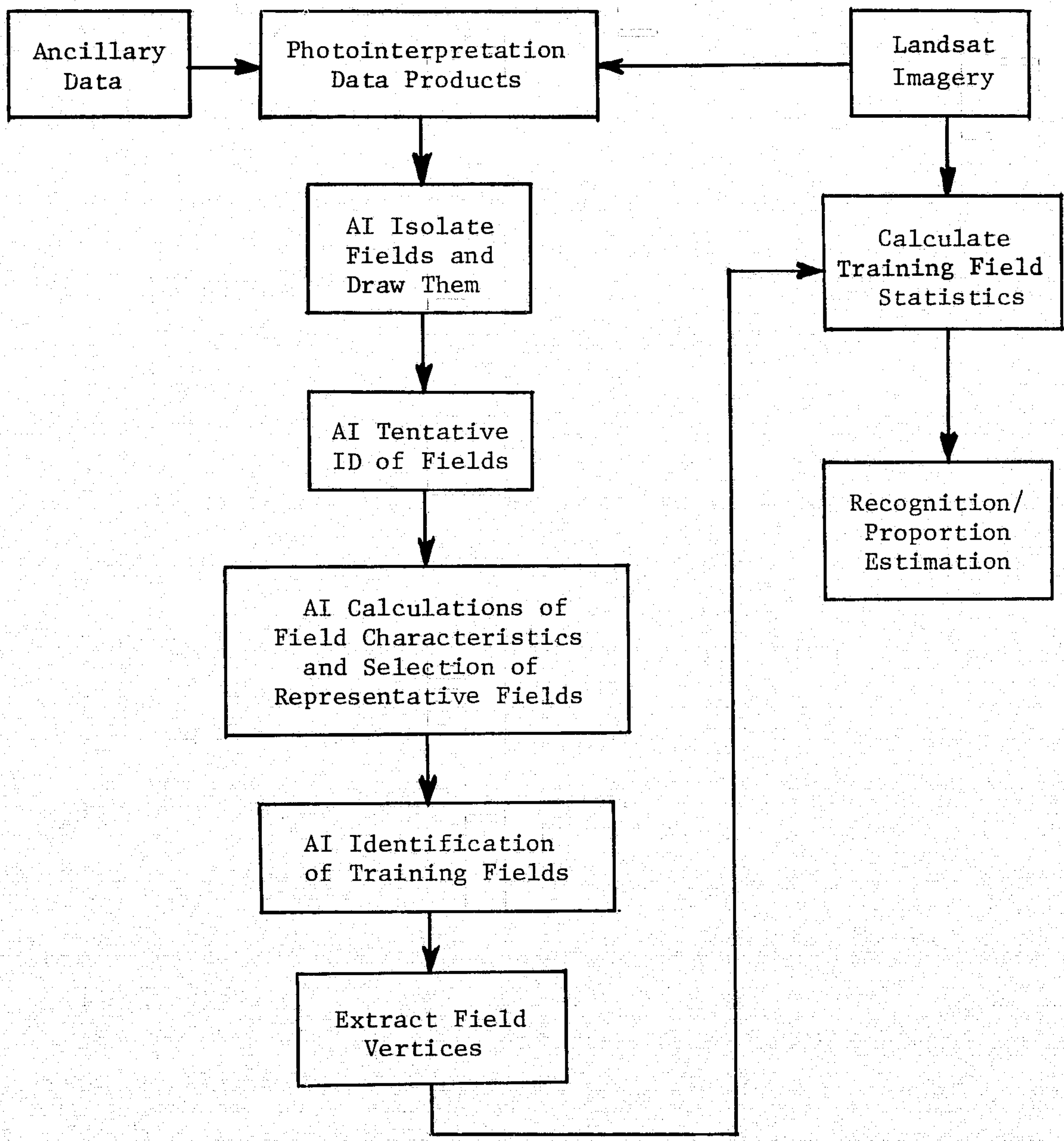


FIGURE 1(a). FUNCTION FLOW, CURRENT CAMS SYSTEM

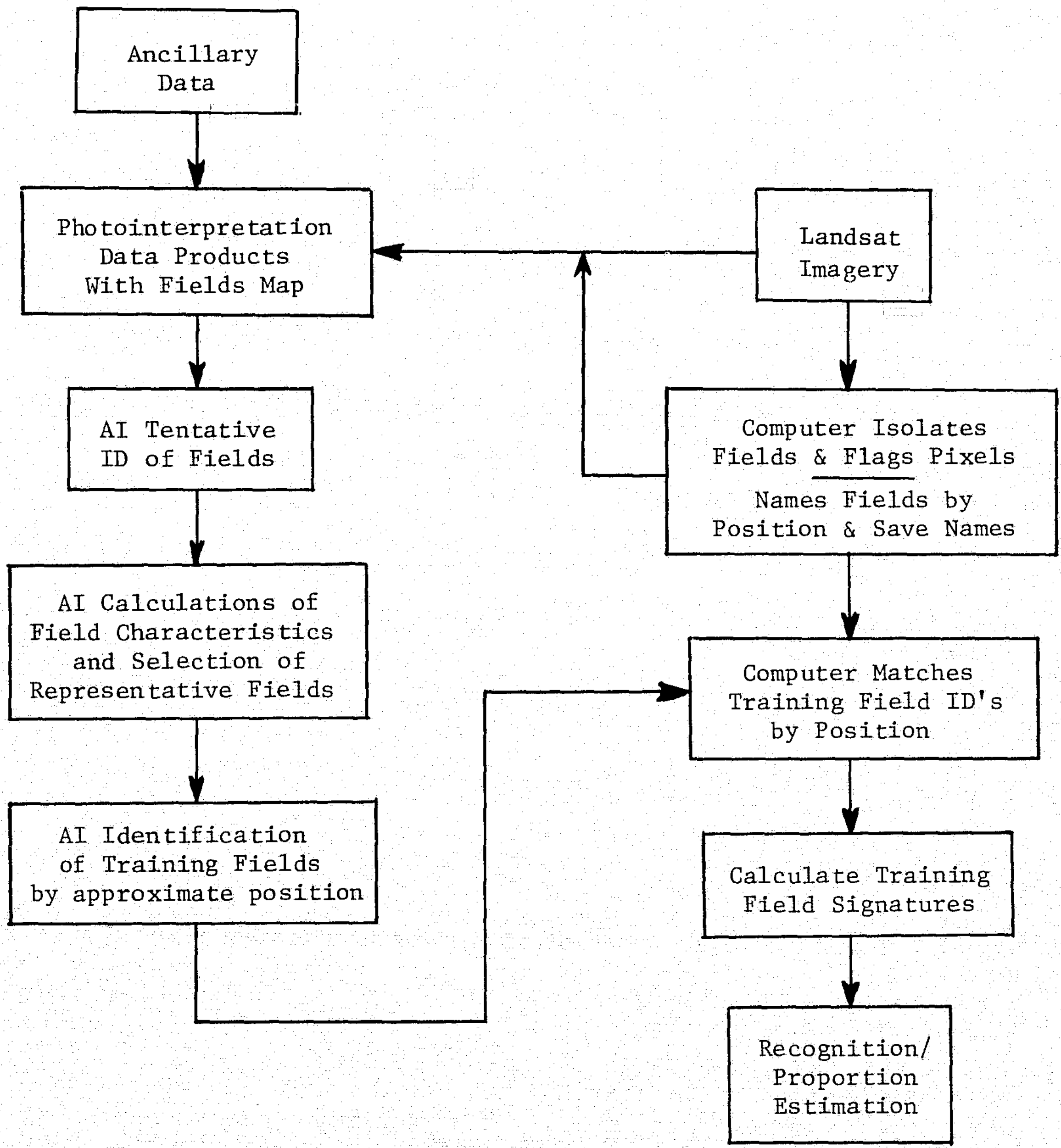


FIGURE 1(b). FUNCTION FLOW WITH PARTIAL COMPUTER SUPPORT TO ISOLATE AND FLAG FIELD PIXELS

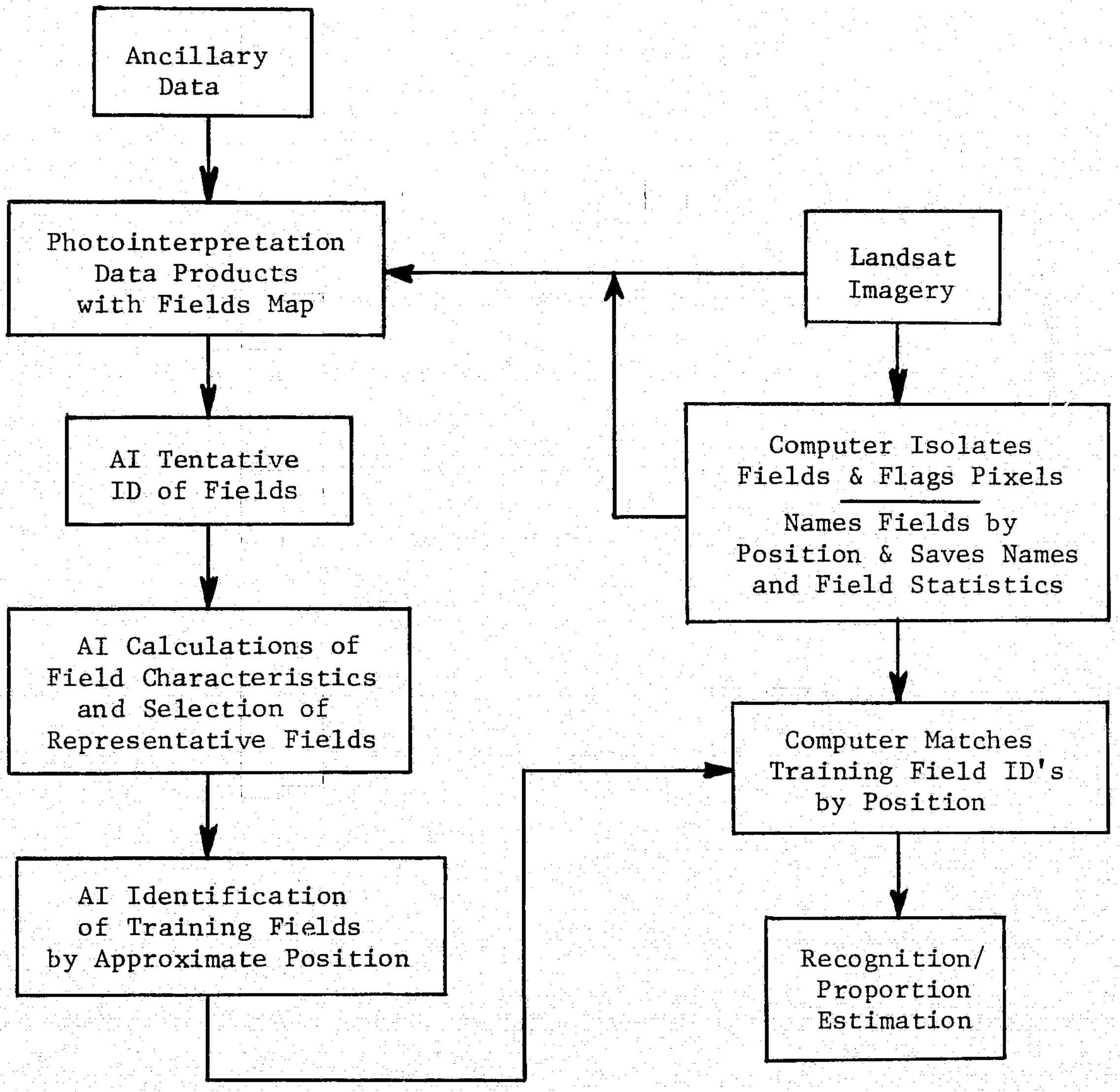


FIGURE 1(c). FUNCTION FLOW WITH PARTIAL COMPUTER SUPPORT TO ISOLATE AND FLAG FIELD PIXELS AND PRECALCULATE STATISTICS

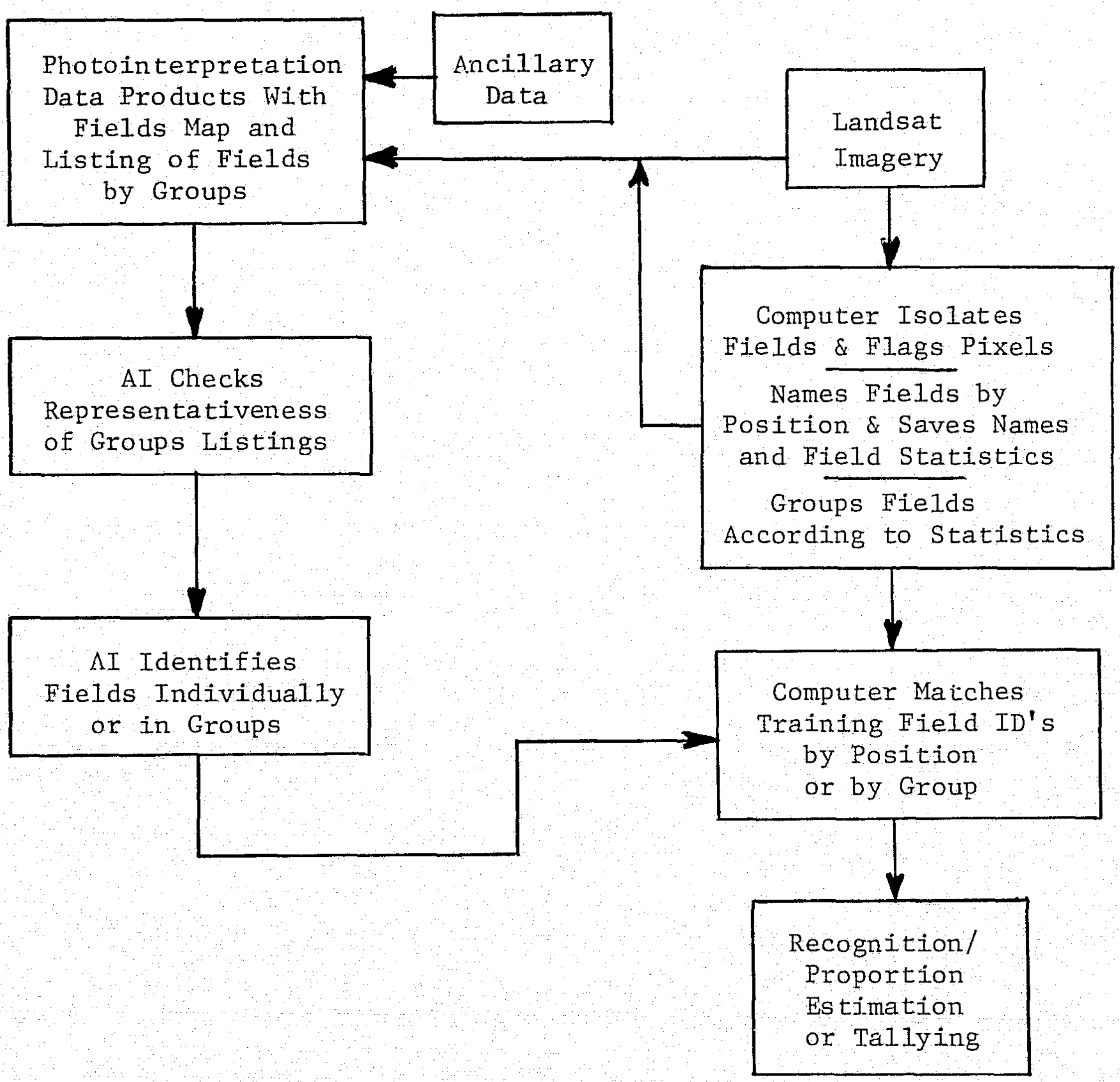


FIGURE 1(d). FUNCTION FLOW WITH PARTIAL COMPUTER SUPPORT TO ISOLATE AND FLAG FIELD PIXELS, PRECALCULATE STATISTICS, AND GROUP FIELDS ACCORDING TO THEIR STATISTICS.

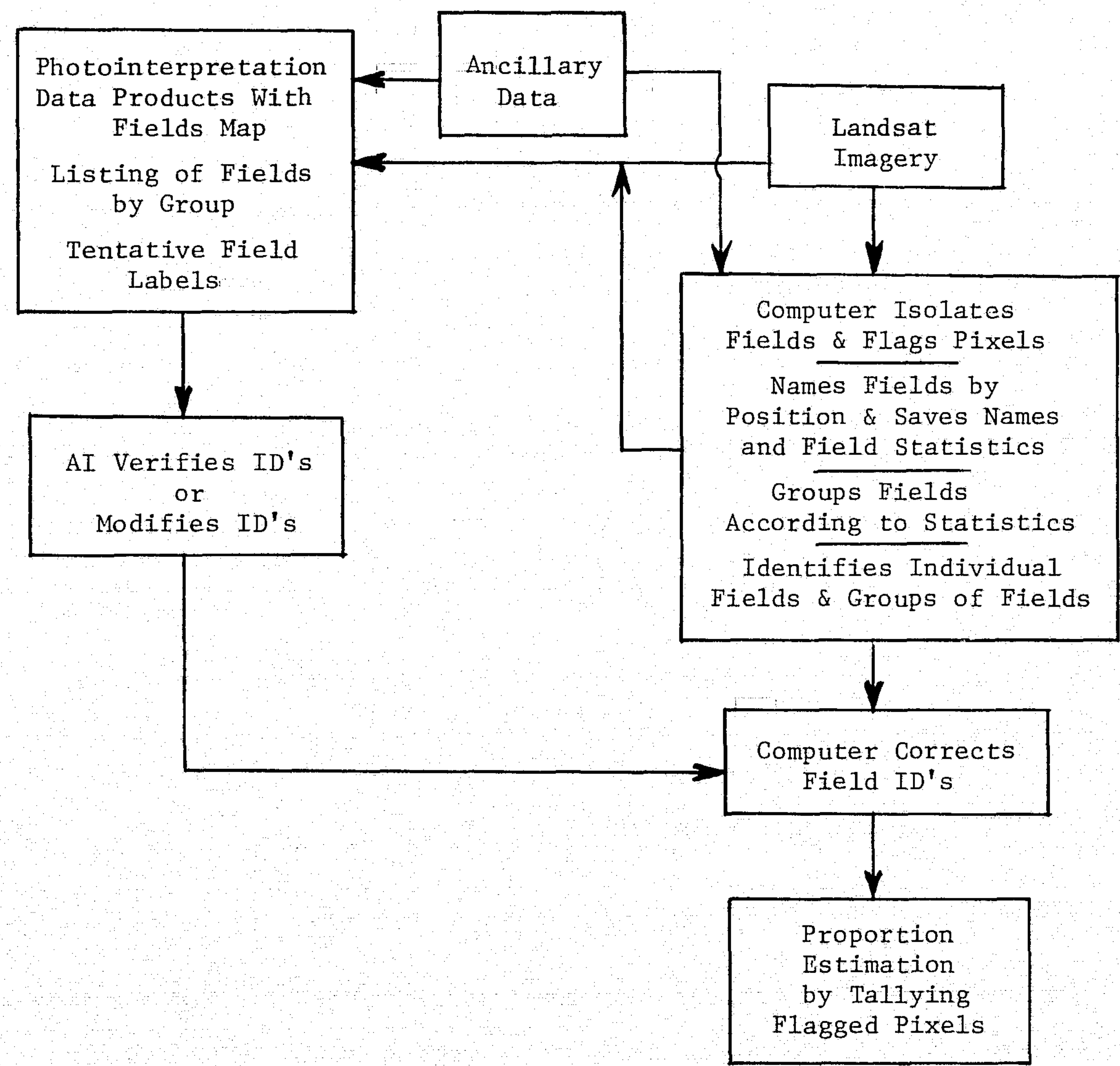


FIGURE 1(e). FUNCTION FLOW WITH COMPUTER SUPPORT TO IDENTIFY FIELDS

training fields, the AI can go through the listing and determine whether he has chosen at least one from each group. The AI could also identify an entire group of fields as being wheat or non-wheat. If all the groups were thus identified, the proportion estimation could be accomplished by tallying the various flagged pixels.

In Figure 1(e), another new computer support function is added, namely the (tentative or final) identification of the fields. In order to carry out this function some external information is required, and so we show ancillary data going to both the computer and to the AI. Actually, the tentative identification of fields is a major increase in complexity, a not surprising fact, since it is the prime function of the AI in the LACIE system. In order to accomplish this function, the computer must, at least implicitly, correct the data for external effects. It must then classify the fields against some multitemporal/multispectral model of crops, whether this model is a data bank or an analytic model. It is pretty clear that crop calendar information will need to be in the model to correct for variations in acquisition time. It appears that the spectral features chosen to enter the model might be similar to the ones the AI uses, namely color and brightness, rather than individual channel responses.

In the above discussion, we see two main classes of computer support function; those which require no ancillary data but which organize the sample segment for the AI in varying degree, and those which assist the AI in identifying fields. The purpose of the first class is either to relieve the AI of some cumbersome detail to make it possible for him to work more sample segments, or to assist him in making certain that his field selections are representative. The ultimate purpose of the second class is to replace most of the AI's in the system. Both of these classes of function are discussed in more detail in later sections.

4
SUPPORTING CAPABILITIES

4.1 FIELD ISOLATION AND ANALYSIS

In the earlier Section 3 several functions were identified in a hierarchy of possible assistance to the AI. These included the isolation of fields, the flagging of field center pixels so that they are associated together and the labeling of field groups of pixels so that the AI can point to them in some way; these are all field isolation steps. Additional steps are the calculation of the signatures of the groups of pixels associated into fields, the collection of fields into spectrally similar groups and the selection of representative fields from those groups.

Several different methods of field isolation were tried and have been mentioned in previous quarterly reports.* These are the method of boundary elimination by gradients, the method of boundary elimination by nine-point rule (Bayes 9 in this case) and the method of field building by "line and point clustering" or "spectral-spatial clustering." In all cases these methods were considered because their implementation involved modest changes to existing computer programs. In addition, some consideration has been given to the field building algorithm of Gupta and Wintz [1], but we have not tried it out, primarily because it would have to be programmed from scratch for our computer systems and we appear to have an acceptable alternate in hand.

The results of the boundary elimination by nine-point rules appear qualitatively to be inferior to the results from spectral-spatial clustering or gradient. Of these latter two, spectral-spatial clustering has the advantage that it automatically incorporates the calculation of the field statistics (with certain reservations which will be explained

* Technical Progress Report, 15 August 1975-14 November 1975, Task 12, Contract NAS9-14123, Report No. 109600-41-L, December 1975.

later) and it automatically designates the field in a way that is potentially useful to the AI, i.e., by the location of the field center. An additional potential advantage is that any clustering program could be modified to do spectral-spatial clustering relatively easily, including ISOCLASS or MLE. Therefore we will proceed to briefly describe the spectral spatial clustering algorithm and present typical results.

The basic idea of clustering algorithms is to group together pixels which are spectrally similar in some sense. The idea of spectral-spatial clustering is to group together pixels which are spectrally and spatially similar in some sense. Normally, each pixel is represented by a vector of spectral intensity values, i.e., Landsat MSS counts. In spectral-spatial clustering these vectors are augmented by the addition of two more components, one for the line number and one for the point number within the line.

The clustering algorithm normally includes or rejects individual pixels from within the clusters based on the probability distribution of the points already included in the cluster. The sample distribution is assumed to be normal with the mean and covariance being the sample mean and the sample covariance of the points already in the cluster. The addition of line and point channels causes each cluster to be localized, to have a mean line and point value and a line and point variance. At the end of the run the cluster statistics are printed. These include the means of line and point which serve well as indicators of the position of the cluster.

In order to avoid a confusion in terms we have followed the example of Gupta and Wintz and have adopted the name "blob" for line and point clusters. This is to distinguish between computer output groups of pixels and true fields, or fields designated by the AI's. In general, spectral-spatial clustering will break "large" fields into more than one blob.

Some additional features of the spectral-spatial clustering program are:

a. Provision is made to place a relative weight on the importance of the spectral channels versus the line and point channels. For multitemporal data the weights are adjusted to place less relative importance on the spectral channels.

b. The raw line and point coordinates are rotated before clustering to produce coordinates which are lined up with a normal map grid. In order to use the means to locate points in LACIE sample segments they must be rotated back again.

The reason the rotation is appropriate is that in order to save time the clustering algorithm uses only line and point variance rather than covariance. The overall effect without the rotation would be that the algorithm would be biased toward producing blobs which are ellipsoidal-shaped but not lined up with the North-South world map. After rotation, the algorithm is biased towards producing blobs which are lined up with a North-South map.

c. A recent modification is that in addition to being rotated, the line and point coordinates use the criteria $C^2 = \sqrt{x^4 + y^4}$ rather than $C^2 = x^2 + y^2$. The effect of this criterion, in the absence of competition from other blobs, is to create a super-ellipsoid shaped blob, which seems to be more field like. In practice neither the rotation nor the super-ellipsoid criterion seem to affect the shape of blobs very much. Instead, blobs crowd together to fill out into corners and the actual shape seems to be determined by competition between neighboring blobs.

d. As mentioned, blobs tend to crowd together to completely fill the sample segment. However, one purpose of this entire exercise is to identify field center pixels. Therefore, the pixels on the boundary between blobs are flagged as boundary pixels. The remaining pixels in each blob constitute "stripped blobs". The stripping of blobs does not significantly alter their mean line and point coordinates.

Figures 2(a), (b), (c), and (d) show portions of LACIE sample segments which have been subjected to multitemporal spectral-spatial clustering. Figure 2(a) shows the Morton Intensive Test Site (ITS). The four dates used are:

23 October 1973
9 May 1974
27 May 1974
14 June 1974.

In Figure 2(a), the ground-truth field lines (and field numbers), and an encompassing rectangle are overlaid on the blob presentation. The field center pixels are left blank while the stripped pixels are shown as asterisks. Some fields are missing entirely, since they were formed into small enough blobs that they were stripped entirely away by the stripping operation. This is no particular drawback since such small or ragged fields aren't likely to make very good candidates for training fields anyway.

A visual survey of Figure 2(a) supports the following statements:

a. Most 160 acre fields provide a potential training field (field 10 and most of its neighbors are 160 acre fields).
b. A majority of 80 acre fields provide a potential training field (field 32 and some of its neighbors are 80 acre fields).
c. A substantial minority of 40 acre fields provide a potential training field.

Note that one line of the data is a bad line, i.e., a line filled with noisy points in one or more channels. Spectrally each point is likely to be extremely different from its neighbors; hence, the line tends to produce a large number of very small blobs confined to the line. The process of stripping then removes the bad line, and the one above and the one below. The line in question is line 248 which passes through the circular field number 144, at the left edge of the figure.

Some large fields are converted into two blobs and so would be used as two separate training fields having the same ID. This effect is

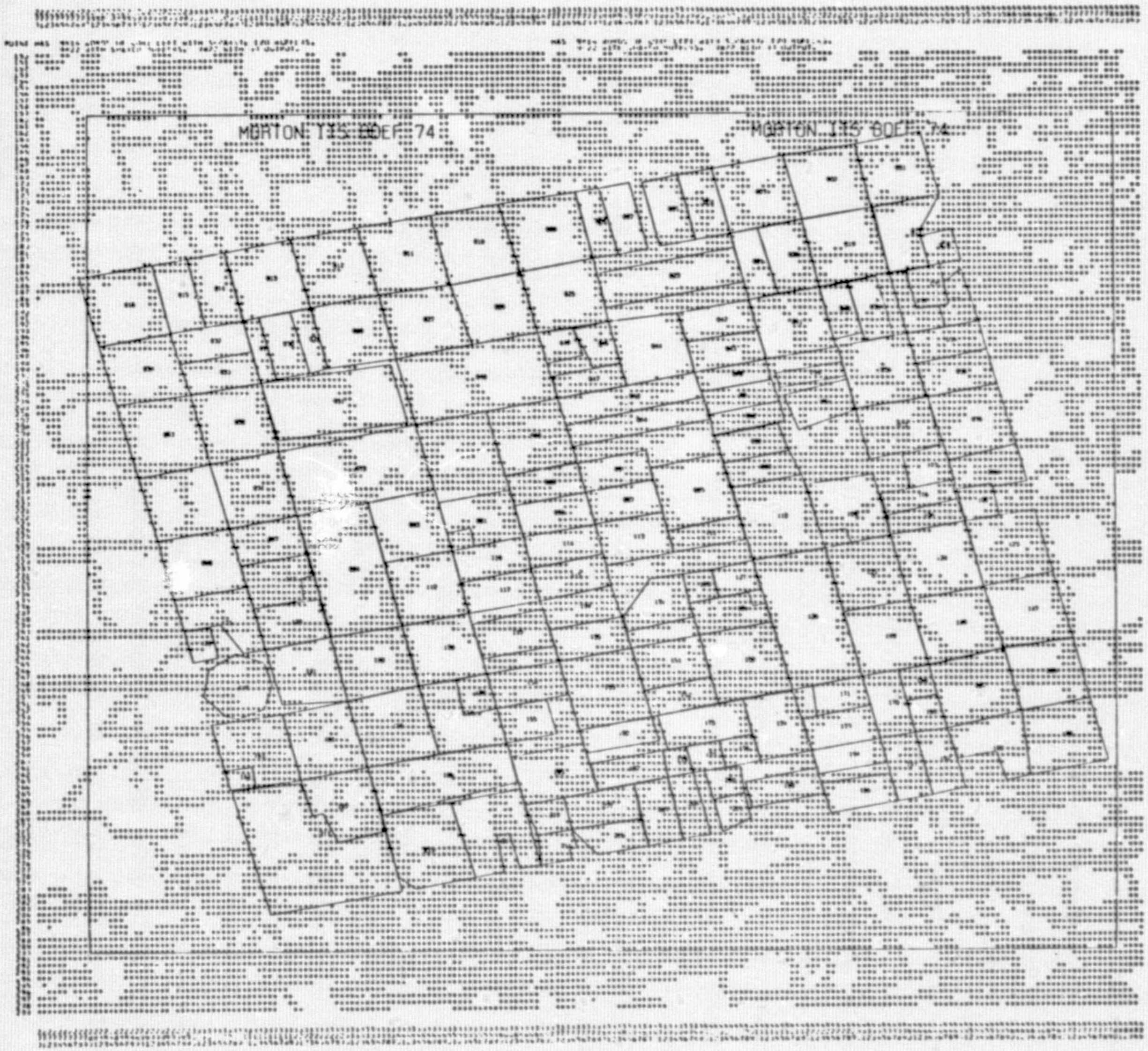


FIGURE 2(a). MORTON ITS MULTITEMPORAL SPECTRAL SPATIAL BLOB MAP

FIGURE 2(b). LANE SRS SITE MULTITEMPORAL SPECTRAL SPATIAL BLOB MAP

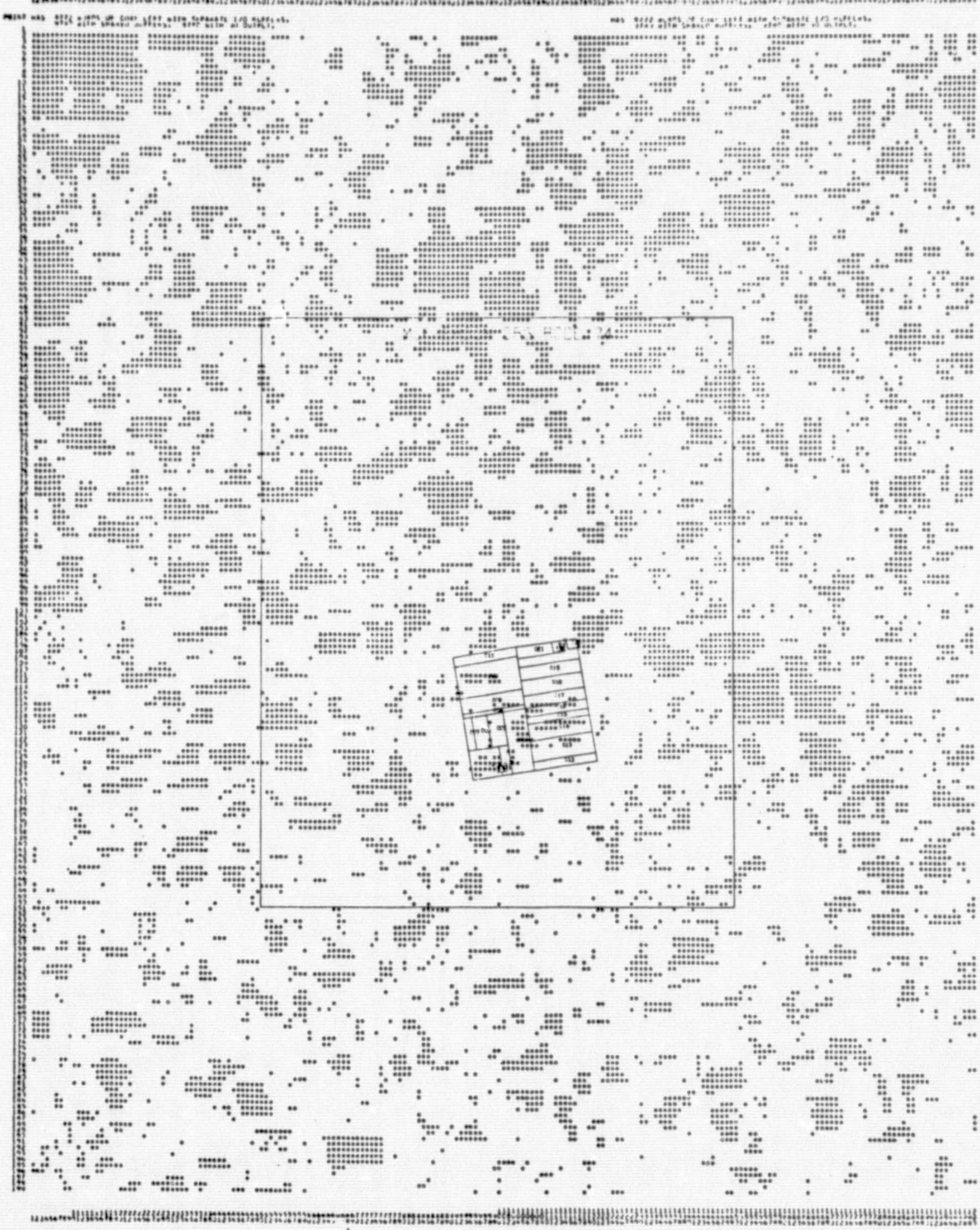


FIGURE 2(c). MCPHERSON SRS SITE MULTITEMPORAL SPECTRAL SPATIAL BLOB MAP

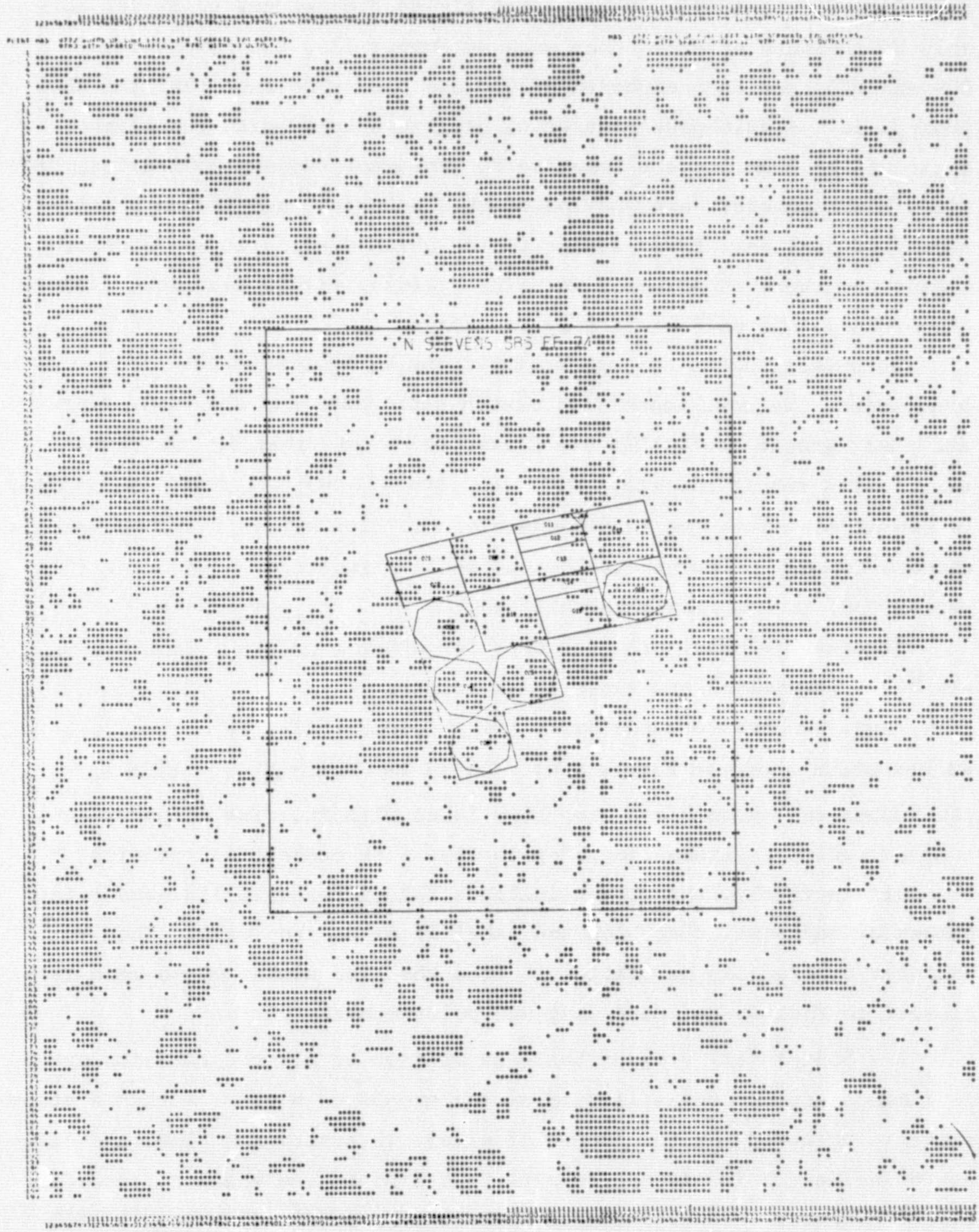


FIGURE 2(d). N. STEVENS SRS SITE MULTITEMPORAL SPECTRAL SPATIAL BLOB MAP

noticeable in the fields which are large in East-West extent, Numbers 049, 051, and 073. Other large fields did not get split up, notably North-South oriented ones such as Number 126, the composite of 009 and 025, and the composite of 010 and 026. The fact that the BLOB program is a single-pass clustering program coupled with the slight angle of the scan pattern relative to East-West running boundaries, provides a satisfactory explanation of this behavior, and it can be corrected if need be. However, the fields which were split are 320-acre fields and there is no point in using training fields that large anyway, since the points will be largely redundant.

In several cases throughout the scene blobs run across field boundaries. We have identified eleven such cases and the field pairs and their ground truth ID's are listed below in Table 2. In 10 out of 13 cases the ID's match. (400, 402 and 404 are all wheat varieties.) In two of the cases the infringement of the blob across field boundaries amounts to only a few pixels, but is still significant. One case (013 versus 030) is completely unexplained.

Figures 2(b)-2(d) are blob maps for other segments. In these maps the blobs themselves are shown as asterisks.

Thus, we have here the basis for a computer AI-interaction such as was shown above in Figure 1(b). The AI can identify the blob by its number and associate a crop type label with it. The computer can then use all the pixels flagged with that blob number for training.

Alternatively, the AI can indicate the approximate line-and-point number of any particular blob and label it as to crop type. Then a computer program can be written to find the blob number whose mean is closest to that approximate line and point number.

A supplementary program has been written to cluster the blob spectral means and produce a listing of the groups of blobs. Within a group they are organized by the number of pixels in a stripped blob, the largest being last. The AI can examine this list to see whether, according to the computer, he has included representatives of all the spectrally distrinct fields in the scene. This input would allow the implementation of Figure 1(c), discussed earlier in Section 3.

TABLE 2. EXAMPLES OF BLOBS OVERLAPPING BOUNDARIES BETWEEN PAIRS OF FIELDS

Field Name	Field ID
002	700
019	700
009	402
025	404
010	700
026	700
013	400
030	700
039	404
040	400
074	700
075	700
073	700
093	400
093	400
094	404
094	404
118	700
118	700
116	700
148	500
149	500
173	700
193	700
185	400
186	400

Table 3 shows the groupings of blobs for the Morton ITS. The number of groups, hereafter called Blob clusters or B clusters, was limited to a maximum of 30 which was thought to be in line with the number of possible spectral classes in a scene.

Under each B cluster are shown the blob numbers associated with that cluster, the number of blob center pixels (unstripped pixels) in each blob and, where the number of unstripped pixels was greater than zero, the line and point means for the blob. Blobs are ranked from smallest to largest. The number of blobs in a B cluster varies. B clusters 1 and 2 contain 52 and 132 blobs respectively. B clusters 27 through 30 contain 1 blob each. The number of unstripped pixels in a blob varies from 0 to 235.

Simulating our proposed AI methodology, we took the ground information, our multitemporal spectral-spatial cluster map, and the B cluster table relevant to Morton and identified B clusters 4 and 10 as all Wheat clusters; B clusters 2, 3, 6, 7, 8, 9, and 19 as containing only "Other"; 4 ambiguous B clusters, numbers 1, 5, 11, and 18, containing a mix of Wheat and Other; and the rest could not be identified from the existing ground truth. (Appendix I contains a listing of field ID's.)

There was at least one instance when some supplementary information was required to explain an apparent inconsistency in the line and point mean of one of the blobs and the very existence of another. In Table 3 near the end of the listing for B cluster 4 you will see blob number 220. Its line and point means fall on a <u>boundary</u> between fields 33 and 52. It was revealed that the blob sitting in <u>field 32</u> was a part of the blob sitting in field 52. The connection between them is visible where it occupies field 31 and is lost at the intersection of fields 31, 33, 51, and 52. Fields 31, 32, and 52 are identified as wheat and field 33 as summer fallow in the ground information for the Morton ITS.

TABLE 3. MORTON ITS BLOB GROUPINGS

****

This table is divided into sections by cluster (B cluster) number. Under each cluster you will find four columns of numbers. In the first column, headed by the word BLOB, is blob number. The second column, headed by NO PTS, contains the number of blob center pixels (unstripped pixels) in the appropriate blob. The third and fourth columns contain the line and point means of blob centers containing at least 1 pixel.

****

CLUSTER NO. 1

BLOB	NO PTS	LINE	POINT
1	0	0	0
86	0	0	0
97	0	0	0
101	0	0	0
166	0	0	0
196	0	0	0
250	0	0	0
265	0	0	0
372	0	0	0
376	0	0	0
382	0	0	0
415	0	0	0
422	0	0	0
424	0	0	0
551	0	0	0
553	0	0	0
595	0	0	0
596	0	0	0
606	0	0	0
639	0	0	0
675	0	0	0
680	0	0	0
706	0	0	0
709	0	0	0
716	0	0	0
778	0	0	0
787	0	0	0
738	0	0	0
625	1	283	352
667	1	298	333
672	1	288	396
685	1	292	353
174	2	179	481
493	2	256	311
544	2	269	411
655	2	288	339
691	2	290	475
705	2	293	301
93	3	159	330
682	3	289	492
543	5	269	390
564	6	277	356
745	6	298	317
580	7	277	377
486	8	256	306
542	8	269	370
177	11	184	472
629	17	294	408
369	22	232	429
698	23	293	427
85	30	162	338
480	41	257	327

TOTAL FOR CLUSTER 1 ARE 206

CLUSTER NO. 2

BLOB	NO PTS	LINE	POINT
2	0	0	0
7	0	0	0
31	0	0	0
33	0	0	0
57	0	0	0
70	0	0	0
79	0	0	0
82	0	0	0
106	0	0	0
125	0	0	0
195	0	0	0
241	0	0	0
248	0	0	0
279	0	0	0
305	0	0	0
311	0	0	0
359	0	0	0
361	0	0	0
366	0	0	0
492	0	0	0
517	0	0	0
521	0	0	0
546	0	0	0
562	0	0	0

TABLE 3. MORTON ITS BLOB GROUPINGS (Continued)

604	0	0	0
661	0	0	0
123	1	166	473
179	1	185	422
107	1	186	454
511	1	263	313
512	1	262	319
536	1	267	376
617	1	282	397
155	2	175	455
191	2	188	450
223	2	200	453
343	2	231	314
531	2	260	332
719	2	293	382
636	3	288	380
779	3	299	443
239	4	202	483
313	4	216	410
145	5	173	483
427	5	250	395
527	5	265	390
619	5	286	417
68	6	155	428
190	6	185	411
494	6	264	383
202	7	193	456
357	7	235	436
426	7	252	365
244	8	201	377
24	9	152	425
302	9	217	311
363	9	232	398
528	9	265	445
335	10	225	433
37	11	152	483
362	11	230	386
392	11	236	308
403	11	257	344
132	12	168	390
577	12	279	484
664	12	287	478
10	14	152	334
22	14	152	415
181	15	184	427
404	15	244	394
406	15	241	308
658	15	288	377
228	18	195	371
141	20	170	364
510	20	267	412
173	22	181	468
3	23	153	274
72	24	157	389
212	25	194	483
267	25	258	391
309	25	217	443
409	27	241	362
273	28	213	353
51	31	154	460
348	31	226	371
402	33	241	376
235	34	200	305
102	35	163	435
278	35	213	379
602	35	286	277
65	36	161	345
142	36	174	376
525	37	269	471
216	38	191	401
468	38	252	436
91	39	162	274
136	39	172	334
158	39	178	303
88	40	160	445
419	40	251	325
368	41	244	322
295	42	214	468
322	46	219	427
122	49	174	392
470	51	257	455
416	52	247	336
242	53	206	451
133	54	169	437
583	54	280	476
98	56	164	465
418	58	247	464
77	62	166	292
474	64	255	413
318	68	222	273
180	75	185	393
293	80	215	324
276	81	212	337
414	82	250	296
161	84	181	412
484	88	258	427
307	90	218	396
317	90	225	472
205	93	192	345
60	97	164	366
168	98	182	341
347	105	229	301
328	113	231	351
413	124	243	282
157	140	179	286
203	141	194	275
111	178	173	424
165	211	185	357

TOTAL FOR CLUSTER 2 ARE 2907
CLUSTER NO. 3

BLOB NO	PTS	LINE	POINT
4	0	0	0
17	0	0	0
27	0	0	0
49	0	0	0
66	0	0	0
73	0	0	0
116	0	0	0
167	0	0	0
211	0	0	0

TABLE 3. MORTON ITS BLOB GROUPINGS (Continued)

213	0	0	0
227	0	0	0
275	0	0	0
277	0	0	0
269	0	0	0
298	0	0	0
329	0	0	0
331	0	0	0
332	0	0	0
337	0	0	0
341	0	0	0
354	0	0	0
366	0	0	0
379	0	0	0
398	0	0	0
490	0	0	0
513	0	0	0
524	0	0	0
538	0	0	0
567	0	0	0
570	0	0	0
582	0	0	0
597	0	0	0
599	0	0	0
622	0	0	0
626	0	0	0
627	0	0	0
635	0	0	0
648	0	0	0
663	0	0	0
686	0	0	0
689	0	0	0
690	0	0	0
739	0	0	0
746	0	0	0
766	0	0	0
790	0	0	0
791	0	0	0
246	1	213	428
284	1	212	445
304	1	214	439
315	1	217	437
475	1	253	460
488	1	255	420
581	1	277	385
23	2	157	428
148	2	176	457
272	2	214	481
408	2	242	340
623	2	289	335
182	3	195	445
290	3	211	459
464	3	250	354
572	3	281	312
247	5	202	473
325	5	218	373
718	5	295	378
208	6	203	444
707	6	293	312
200	7	196	455
389	7	238	404
515	7	266	383
353	8	232	441
561	8	274	372
575	8	277	422
683	8	294	271
262	9	206	405
126	10	168	271
306	12	237	301
603	14	284	302
8	15	153	312
550	16	272	382
552	16	272	419
147	19	180	456
127	21	171	321
342	22	225	315
498	23	263	446
565	23	276	401
558	25	273	476
485	26	255	469
44	29	156	317
39	30	162	318
206	31	217	314
500	33	260	465
319	40	220	295
549	46	273	345
49	47	155	404
146	59	163	303
285	62	220	482
306	63	218	303
69	74	166	450
476	94	256	481
371	179	239	419

TOTAL FOR CLUSTER 3 ARE 1151

CLUSTER NO. 4

BLOB	NO PTS	LINE	POINT
6	0	0	0
13	0	0	0
20	0	0	0
21	0	0	0
29	0	0	0
41	0	0	0
50	0	0	0
52	0	0	0
99	0	0	0
100	0	0	0
150	0	0	0
151	0	0	0
215	0	0	0
283	0	0	0
351	0	0	0
352	0	0	0
428	0	0	0
463	0	0	0
616	0	0	0
651	0	0	0
669	0	0	0
711	0	0	0
740	0	0	0
90	1	158	472

TABLE 3. MORTON ITS BLOB GROUPINGS (Continued)

495	1	257	389
533	1	278	396
30	2	152	444
43	2	152	478
61	2	155	436
89	2	158	453
163	2	181	482
207	2	195	448
460	2	250	439
18	3	155	382
80	3	157	481
162	3	176	462
420	3	245	365
514	3	265	338
768	3	298	450
25	4	153	432
310	4	216	451
529	4	265	458
206	5	190	416
412	5	241	483
237	6	208	368
560	6	275	364
5	7	152	286
300	7	215	283
28	8	154	446
170	8	179	397
384	8	233	370
19	9	152	388
104	9	169	270
754	9	297	357
53	11	162	481
9	13	164	327
103	13	160	474
472	13	252	271
522	13	264	416
140	15	173	465
425	16	257	299
224	17	196	323
421	17	248	310
63	19	164	325
14	21	154	375
286	24	212	363
467	27	250	424
621	28	284	450
240	29	205	378
479	30	253	367
71	33	161	352
508	33	262	347
538	33	281	466
423	35	251	451
67	36	158	360
198	37	189	429
333	37	222	327
266	38	212	272
222	39	194	413
231	40	199	430
55	41	158	295
137	42	173	402
287	42	219	385
95	44	160	379
87	45	171	345
523	46	266	430
176	49	184	330
130	50	171	354
324	53	221	352
186	54	200	461
502	54	262	364
355	55	235	358
251	58	207	466
537	59	270	457
189	63	188	297
741	63	297	396
153	68	178	438
118	69	169	306
131	74	173	383
236	81	202	347
400	82	243	345
76	83	165	282
175	84	190	286
373	86	237	462
96	87	162	395
105	87	168	411
140	87	174	324
320	91	223	340
554	93	275	434
292	95	221	285
233	98	200	361
367	113	234	275
374	122	235	478
38	123	157	304
178	125	195	389
360	129	232	288
340	132	231	332
220	142	202	307
160	160	187	316
152	195	193	371

TOTAL FOR CLUSTER 4 ARE 4124
CLUSTER NO. 5

BLOB	NO PTS	LINE	POINT
12	0	0	0
15	0	0	0
16	0	0	0
40	0	0	0
45	0	0	0
46	0	0	0
47	0	0	0
54	0	0	0
59	0	0	0
64	0	0	0
75	0	0	0
108	0	0	0
164	0	0	0
167	0	0	0
226	0	0	0
253	0	0	0
255	0	0	0
257	0	0	0
301	0	0	0
303	0	0	0
312	0	0	0

TABLE 3. MORTON ITS BLOB GROUPINGS (Continued)

350	0	0	0
479	0	0	0
503	0	0	0
507	0	0	0
520	0	0	0
534	0	0	0
566	0	0	0
585	0	0	0
640	0	0	0
659	0	0	0
660	0	0	0
730	0	0	0
743	0	0	0
747	0	0	0
749	0	0	0
783	0	0	0
260	1	213	477
316	1	216	449
335	1	230	409
345	1	226	395
364	1	229	419
377	1	234	384
487	1	255	374
593	1	279	414
594	1	283	344
643	1	285	345
11	2	152	344
26	2	152	436
145	2	168	465
219	2	190	454
225	2	192	440
399	2	241	383
557	2	274	452
607	2	286	400
210	3	189	463
385	3	234	396
417	3	244	455
612	3	282	275
630	3	295	427
516	5	263	308
708	5	294	332
134	6	168	460
143	6	170	482
238	6	199	416
339	6	226	460
634	6	285	341
274	7	209	427
532	8	268	363
547	9	276	443
548	9	279	337
62	11	154	473
171	11	180	444
777	11	249	412
109	12	162	421
258	12	207	475
256	13	211	406
568	13	277	391
229	15	195	378
491	15	249	379
482	15	257	376
632	17	288	313
330	18	227	376
613	18	282	324
243	19	202	276
530	19	278	319
657	19	288	367
327	21	224	455
264	22	207	436
234	23	212	431
601	26	234	435
334	27	222	367
633	28	286	324
631	30	285	291
281	31	211	396
411	31	244	477
559	32	275	331
681	34	290	440
701	42	294	282
699	43	292	466
611	45	285	461
556	49	274	409
410	50	244	406
507	52	262	299
199	53	197	435
144	54	181	272
688	55	294	411
344	74	229	387
346	83	228	447
214	84	193	330
221	85	202	336
535	91	270	302
249	93	205	322
356	100	242	435
526	112	270	317
477	128	265	286
462	168	257	276
545	235	275	281

TCTAL FCR CLUSTER 5 ARE 2353

CLUSTER NC. 6

BLOB NO	PTS	LINE	POINT
32	0	0	0
321	0	0	0
391	0	0	0
465	1	250	361
491	5	260	437
539	6	269	403
403	19	241	394
489	19	269	441

TCTAL FCR CLUSTER 6 ARE 50

CLUSTER NC. 7

BLOB NO	PTS	LINE	POINT
34	0	0	0
81	0	0	0
92	0	0	0
94	0	0	0
112	0	0	0
113	0	0	0
128	0	0	0
154	0	0	0
192	0	0	0

ORIGINAL PAGE IS
OF POOR QUALITY

TABLE 3. MORTON ITS BLOB GROUPINGS (Continued)

BLOB NO	PTS	LINE	POINT
340	0	0	0
370	0	0	0
407	0	0	0
417	0	0	0
646	0	0	0
657	0	0	0
656	0	0	0
702	0	0	0
713	0	0	0
751	0	0	0
764	0	0	0
772	0	0	0
589	1	279	300
473	65	254	398
TOTAL FOR CLUSTER 7 ARE 66			
CLUSTER NO. 8			
BLOB NO	PTS	LINE	POINT
35	0	0	0
36	0	0	0
138	0	0	0
194	0	0	0
204	0	0	0
230	0	0	0
605	0	0	0
620	0	0	0
666	0	0	0
673	0	0	0
679	0	0	0
728	0	0	0
781	0	0	0
387	4	239	316
291	10	211	483
188	63	192	475
TOTAL FOR CLUSTER 8 ARE 77			
CLUSTER NO. 9			
BLOB NO	PTS	LINE	POINT
42	0	0	0
74	0	0	0
73	0	0	0
187	0	0	0
218	0	0	0
294	0	0	0
390	0	0	0
393	0	0	0
618	0	0	0
628	0	0	0
652	0	0	0
654	0	0	0
662	0	0	0
712	0	0	0
721	0	0	0
724	0	0	0
786	0	0	0
789	0	0	0
471	6	258	476
540	8	269	483
124	10	167	477
569	10	275	460
378	13	232	405
297	14	221	452
394	19	237	320
245	24	201	406
TOTAL FOR CLUSTER 9 ARE 104			
CLUSTER NO. 10			
BLOB NO	PTS	LINE	POINT
56	0	0	0
107	0	0	0
110	0	0	0
259	0	0	0
269	0	0	0
314	0	0	0
326	0	0	0
744	0	0	0
280	1	211	433
TOTAL FOR CLUSTER 10 ARE 1			
CLUSTER NO. 11			
BLOB NO	PTS	LINE	POINT
58	0	0	0
323	0	0	0
261	1	206	287
574	1	276	388
647	1	286	432
769	1	299	458
271	2	207	479
732	2	295	444
139	4	170	477
687	4	290	363
280	6	209	442
397	8	237	374
358	9	227	461
270	10	237	422
263	14	209	418
734	15	295	483
499	20	261	401
252	40	227	278
591	51	282	368
308	61	222	414
395	74	239	448
254	90	200	295
TOTAL FOR CLUSTER 11 ARE 414			
CLUSTER NO. 12			
BLOB NO	PTS	LINE	POINT
792	0	0	0
TOTAL FOR CLUSTER 12 ARE 0			
CLUSTER NO. 13			
BLOB NO	PTS	LINE	POINT
84	0	0	0
185	0	0	0
282	0	0	0
586	0	0	0
624	0	0	0
640	0	0	0
653	0	0	0
676	0	0	0
725	0	0	0
737	0	0	0
753	0	0	0
773	0	0	0
774	0	0	0
695	15	293	420

TABLE 3. MORTON ITS BLOB GROUPINGS (Continued)

742	17	296	440

TOTAL FOR CLUSTER 13 ARE 32

CLUSTER NO. 14

BLOB NO	PTS	LINE	POINT
114	0	0	0
117	0	0	0
520	0	0	0
673	0	0	0
693	0	0	0
714	0	0	0
727	0	0	0
752	0	0	0
756	0	0	0
782	0	0	0
684	3	292	354
615	5	283	360
672	9	290	296
645	14	286	350

TOTAL FOR CLUSTER 14 ARE 31

CLUSTER NO. 15

BLOB NO	PTS	LINE	POINT
115	0	0	0
121	0	0	0
763	0	0	0

TOTAL FOR CLUSTER 15 ARE 0

CLUSTER NO. 16

BLOB NO	PTS	LINE	POINT
117	0	0	0
268	0	0	0
405	0	0	0
430	0	0	0
433	0	0	0
436	0	0	0
442	0	0	0
447	0	0	0
449	0	0	0
454	0	0	0
456	0	0	0
506	0	0	0
697	0	0	0

TOTAL FOR CLUSTER 16 ARE 0

CLUSTER NO. 17

BLOB NO	PTS	LINE	POINT
120	0	0	0
129	0	0	0
704	0	0	0
735	1	299	296
715	2	294	296

TOTAL FOR CLUSTER 17 ARE 3

CLUSTER NO. 18

BLOB NO	PTS	LINE	POINT
156	0	0	0
159	0	0	0
338	0	0	0
567	0	0	0
637	0	0	0
642	0	0	0
668	0	0	0
729	0	0	0
755	0	0	0
767	0	0	0
785	0	0	0
496	2	258	292
617	3	281	447
399	4	238	409
731	4	294	367
720	5	293	390
722	5	293	457
629	8	283	384
608	12	292	390
753	12	299	344
692	13	295	343
217	14	193	422
518	18	265	325
638	22	288	451
519	25	267	331
717	26	298	376
736	33	297	323
531	47	269	354
466	55	257	383
481	58	255	353
232	72	199	291

TOTAL FOR CLUSTER 18 ARE 438

CLUSTER NO. 19

BLOB NO	PTS	LINE	POINT
172	0	0	0
184	0	0	0
299	0	0	0
365	0	0	0
571	0	0	0
575	0	0	0
578	0	0	0
600	0	0	0
614	0	0	0
665	0	0	0
674	0	0	0
694	0	0	0
700	0	0	0
726	0	0	0
750	0	0	0
771	0	0	0
775	0	0	0
784	0	0	0
644	2	290	341
565	5	272	377
573	13	276	346
579	14	278	307

TOTAL FOR CLUSTER 19 ARE 34

CLUSTER NO. 20

BLOB NO	PTS	LINE	POINT
183	0	0	0
598	0	0	0
671	0	0	0
677	0	0	0
748	0	0	0
757	0	0	0
758	0	0	0

TOTAL FOR CLUSTER 20 ARE 0

CLUSTER NO. 21

BLOB NO	PTS	LINE	POINT

ORIGINAL PAGE IS OF POOR QUALITY

TABLE 3. MORTON ITS BLOB GROUPINGS (Concluded)

```
                 193        0        0        0
                 380        0        0        0
                 703        0        0        0
                 776        0        0        0
                 760        0        0        0
TOTAL FOR CLUSTER    21 ARE          0
CLUSTER NO.     22
               BLOB NO PTS     LINE   POINT
                 200        0        0        0
                 201        0        0        0
                 381        0        0        0
                 761        0        0        0
                 765        0        0        0
TOTAL FOR CLUSTER    22 ARE          0
CLUSTER NO.     23
               BLOB NO PTS     LINE   POINT
                 375        0        0        0
TOTAL FOR CLUSTER    23 ARE          0
CLUSTER NO.     24
               BLOB NO PTS     LINE   POINT
                 363        0        0        0
                 695        0        0        0
                 710        0        0        0
TOTAL FOR CLUSTER    24 ARE          0
CLUSTER NO.     25
               BLOB NO PTS     LINE   POINT
                 429        0        0        0
                 434        0        0        0
                 435        0        0        0
                 437        0        0        0
                 438        0        0        0
                 441        0        0        0
                 443        0        0        0
                 444        0        0        0
                 445        0        0        0
                 446        0        0        0
                 448        0        0        0
                 450        0        0        0
                 453        0        0        0
                 460        0        0        0
                 461        0        0        0
TOTAL FOR CLUSTER    25 ARE          0
CLUSTER NO.     26
               BLOB NO PTS     LINE   POINT
                 431        0        0        0
                 432        0        0        0
                 439        0        0        0
                 440        0        0        0
                 451        0        0        0
                 457        0        0        0
                 459        0        0        0
                 505        0        0        0
TOTAL FOR CLUSTER    26 ARE          0
CLUSTER NO.     27
               BLOB NO PTS     LINE   POINT
                 452        0        0        0
TOTAL FOR CLUSTER    27 ARE          0
CLUSTER NO.     28
               BLOB NO PTS     LINE   POINT
                 455        0        0        0
                 504        0        0        0
TOTAL FOR CLUSTER    28 ARE          0
CLUSTER NO.     29
               BLOB NO PTS     LINE   POINT
                 458        0        0        0
TOTAL FOR CLUSTER    29 ARE          0
CLUSTER NO.     30
               BLOB NO PTS     LINE   POINT
                 541        0        0        0
TOTAL FOR CLUSTER    30 ARE          0
```

From another point of view, if one were attempting to find the blob number associated with field number 32, one would search the table in vain to find a pair of line and point means which fall in field 32.

Figures 2(b), (c), and (d) present an assortment of blob maps in which the blob centers are filled in and edges blanked out. At the top of each encompassing rectangle we show the site name, a code for the dates of coverage used in generating these maps and two digits showing the year of coverage. The data code is described below:

Letter Code	Time Frame of Data Collection
A	4 October 1973
B	20-23 October 1973
C	18-20 April 1974
D	6-9 May 1974
E	24-27 May 1974
F	12-14 June 1974

For the Lane SRS site time periods A, C, and E were used, for McPherson, B, C, D, and E, and for North Stevens, E, and F. Clustering was limited to Landsat Bands 5 and 6 from each utilized time period. This was done to limit the amount of time spent in the clustering process.

A number of observations may be made upon viewing the maps:

1. Data sites vary considerably in the spatial texture of their cluster patterns. This spatial texture is seen to be the most orderly (blobs of goodly size, regularly distributed) in the Lane site and the least orderly in the McPherson site. Variations in orderliness within a map may also be seen.
2. Road patterns (as field boundaries) can be easily spotted in the Lane and North Stevens maps. One might not know where to drive looking at the McPherson map. Also, more field edges are aligned in the top of map to bottom of map direction (rather than the typical Landsat slant direction) in the McPherson site than in the other two sites.

3. Blob size is in most cases consistent with field size. If the blobs were allowed to grow any further, the spatial information provided by field edges could be lost.
4. Bad data lines were treated specially. This was necessary to reduce their impact on the maps. The technique used to mark the boundaries of fields would have made a triplet of bad lines from each single one. Now, a single bad line shows up as a single line of blanks in the maps.
5. Data consisting of fill bits (used to occupy the data space beyond the edge of a Landsat frame) is flagged and shows as a region of blanks (see the Lane map).

Going beyond observation, we then attempted, for all 4 sites, to establish bounds on the probable wheat content for each site. Our procedure follows:

Step 1. Using the tables of clustered blobs (B clusters) and ground information from a site, establish the identity of as many of the blobs as possible starting with the largest in any B cluster.

Step 2. Establish three classes of B clusters -- known and unique (either all Wheat or all Other), known and ambiguous (some Wheat blobs and some Other blobs) and unknown -- based on Step 1.

Step 3. Tally the total number of pixels in a site and the number of pixels in known and unique B clusters.

Step 4. Establish a lower wheat bound by dividing the number of pixels in known and unique B clusters assigned to wheat by the total number of pixels in the site.

Step 5. Establish an upper wheat bound by dividing the total number of pixels in a scene minus the number of pixels in known and unique B clusters assigned to other by the total number of pixels in a scene.

Step 6. Compute a single point estimate of wheat for a site based on a division of the number of pixels in known and unique clusters assigned to wheat by the sum of those pixels and the number of pixels in known and unique clusters assigned to "Other".

Step 7. Express bounds and estimates as a percentage of the scene.

This methodology assures us of achieving the lowest lower limit and the highest upper limit for wheat because in the calculation for the lower limit all ambiguous and unknown pixels are weighted with the Non-Wheat category whereas in the calculation of the upper limit those ambiguous and unknown pixels are weighted with the Wheat category.

For Morton the site is represented by a rectangle which surrounds the ITS and utilizes the ITS corners as its own. For the Lane, McPherson and North Stevens sites statistics were computed for two areas each: the first area is a near perfect fit to the outline of the SRS site; the second is an expansion around the SRS site including lines 41 through 157 and point numbers 1 through 196. This second area represents an SRS site reprogrammed as an ITS.* Figure 3 shows the results of the wheat estimation. The crossbar on the line connecting the lower and upper bounds for wheat percentage represents the single point estimate from Step 6. Actual percentages of wheat for the Morton ITS and the three SRS sites are shown as asterisks.

Upper and lower bounds on the wheat estimate for Morton were 54.7% and 27.6% respectively; the single point estimate was 37.9%, and; the actual percentage based on ground information was 40.4%. As with Morton the Lane SRS site results were in close agreement with ground information. Upper bound, lower bound and single point estimates for the Lane SRS site were 30.2%, 27.2% and 28.0% respectively (actual

* This definition corresponds to the one being used by RT&E Branch to establish expanded ground truth for these sites.

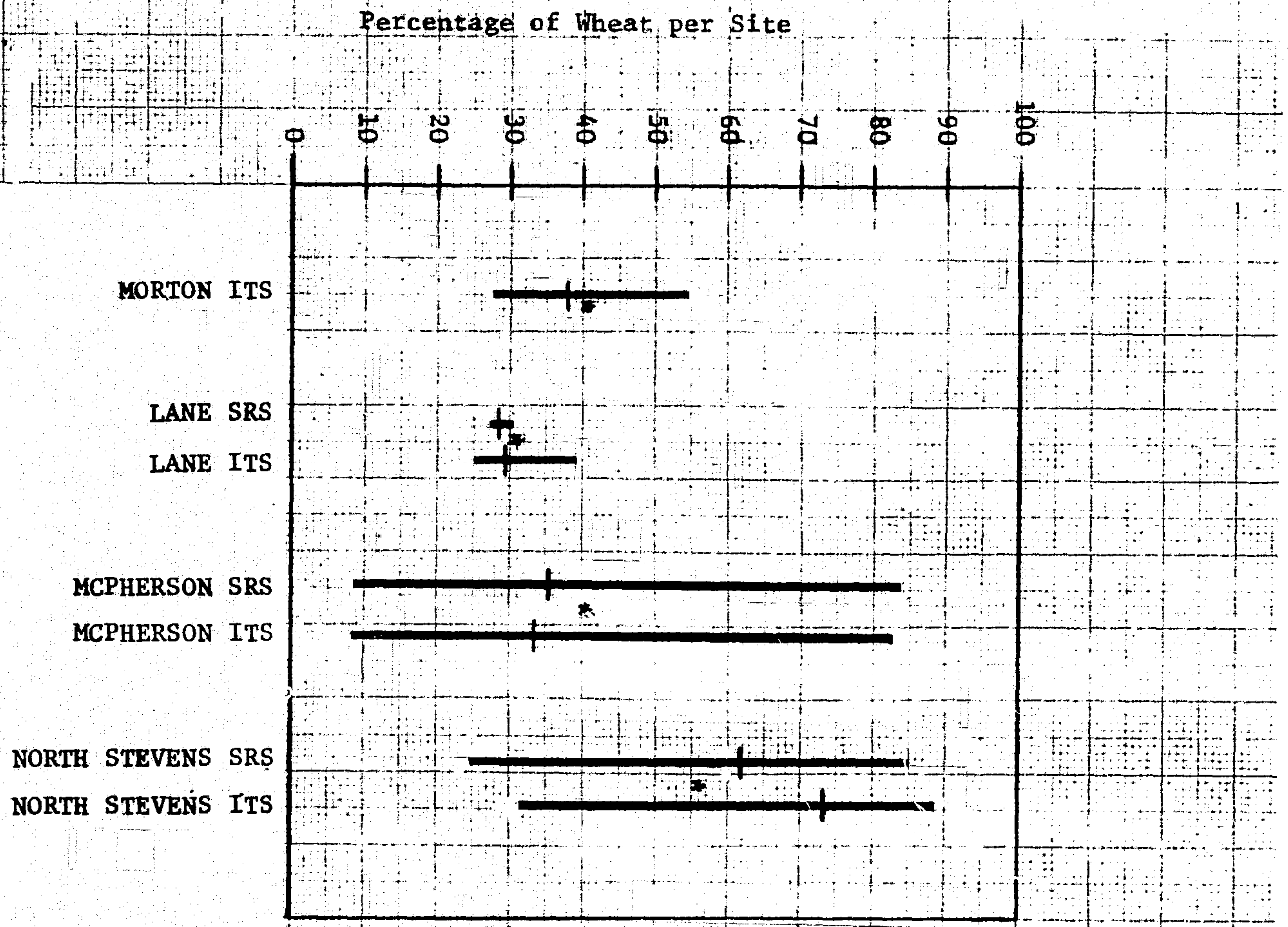


FIGURE 3. COMPARISON OF WHEAT ESTIMATE BOUNDS FOR FOUR SITES

percentage 30.7). In the remaining two sites uncertainty in the wheat estimate increased, as indicated by the spread of the bounds of the estimate. Lower and upper bounds for the McPherson SRS site were 8.9% and 83.7% and for the North Stevens SRS site 24.4% and 84.9%. These quantitative measures seem to follow the qualitative estimates (spatial texture in particular) presented earlier. In spite of the large separation of the wheat estimate bounds for the McPherson and North Stevens SRS sites their single point estimates of 35.2% and 61.8% are reasonably close to their actual wheat percentages of 40.6 and 55.9.

In three out of four cases we underestimated the amount of wheat at a site, coming within about 5% (in an absolute sense). In one case we overestimated by about 6%. Figures for actual wheat in the expanded SRS sites are not yet available. Hence, our estimates for the expanded sites may be regarded, for now, as predictions.

In a more utilitarian vein we must ask how useful blob maps and associated tables would be to the AIs. Granted that the AI could nearly perform a classification on a data set (a ready means of tallying pixels per cluster in a site is needed, among other things) we are uncertain as to whether some of the materials might be more effective in other formats. As an example, the blob maps could be produced to the same scale as the Product One imagery the AIs use and aid in associating blobs on one with color patches on the other. In order to obtain some near future feedback on these questions, samples of available computer aids (blob maps and so on) should be provided to the AIs for evaluation and commentary.

Future work should be concerned with improving the materials AIs can use. Under consideration are B cluster maps which would look like the blob maps for the three SRS sites discussed in this report except that instead of the blob centers being filled with asterisks they would be filled with the number of the B cluster into which the appropriate blob was grouped. In addition, some means should be found by

which the AIs can tally pixels assigned to specific B clusters. AIs should also have the capability to recluster blobs assigned to ambiguous B clusters. Reducing the number of blobs assigned to ambiguous B clusters will have the effect of reducing the difference between the wheat upper bound and wheat lower bound as we have discussed them and hence reduce our uncertainty about any single point estimate of wheat proportion at a site.

4.2 FIELD IDENTIFICATION

The next step in the computer aided ID of wheat is for the computer to, at least tentatively, suggest to the AI the field ID. All steps previous to this point have involved only the data within the sample segment without reference to the world of other sample segments. To this point the only link to the outside is the AI himself, his experience and his training. In order for the computer to say anything about the crop identity it must also bring in outside information. Basically this information will be some expectation of the characteristics of the wheat signals, i.e., a multitemporal-multispectral signature for wheat.

Figure 4 is an expansion of Figure 1(e) with emphasis on computer functions. It shows the struture of a complete system for computer assisted local recognition. Certain of the elements are self explanatory, such as the bad line, cloud and cloud shadow flagging. For others the general intent is clear but the reader, at this point, may have no clear idea as to where the overall concepts are coming from. In particular, the haze correction, non-linear feature extraction, and signature model are concepts all related to each other. There is a nexus of concepts and ideas which need to be explained all together before any of them make sense alone.

4.2.1 GENERAL DISCUSSION OF LANDSAT DATA STRUCTURE

We will start by talking about the gross spectral structure of Landsat data from an agricultural region. Empirical and model ele-

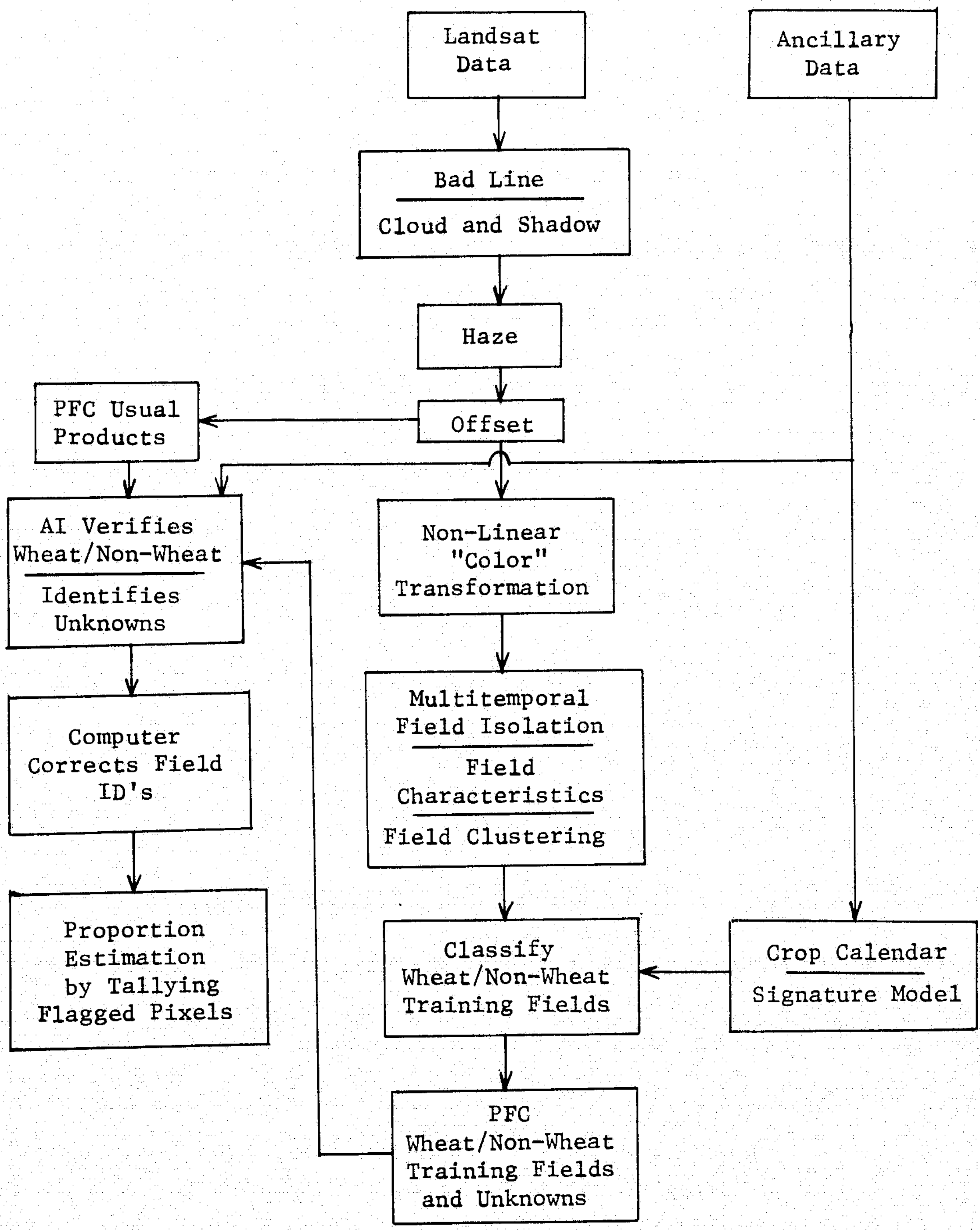


FIGURE 4. OPERATIONS REQUIRED FOR COMPUTER ASSISTED IDENTIFICATION OF WHEAT TRAINING FIELDS

ments are combined in a heuristic idea, the Tasselled Cap. We then examine in more detail some of the implications for processing of Landsat data to extract information from agricultural scenes, and in particular, to identify wheat clusters.

Figure 5(a) shows a two channel scatter diagram of Landsat data in an agricultural scene in Fayette Co., Illinois. The data has been compressed by unsupervised spectral clustering of the data points in all 4 Landsat MSS channels. The ellipses shown are the unit contour ellipses of the normal density function describing each cluster. Shown with each ellipse is an arbitrary number followed by the percent of the scene represented by the cluster generating the ellipse. The channels shown are CH 2 and CH 3 (Landsat Bands 5 and 6).

Notice in Figure 5(a) the definite boundary region near the diagonal of the two channel presentation. All of the agricultural data lies to the left of this boundary. To the right of the boundary there is no data. The region to the left shows a definite triangle like shape, with two vertices on the diagonal and one near the CH 3 axis.

Figure 5(b) shows a similar cluster plot of CH 1 vs. CH 2. Here, all of the data lies near a diagonal of the space again. Thus, we can infer that the triangle-shaped region of Fig. 5(a) is shown edge-on in Fig. 5(b). The three-dimensional shape of the data structure is that of a flattened triangle shape having little thickness.

Figure 5(c) shows a cluster plot of CH 3 vs. CH 4. Again the data lies closely along a diagonal. Viewing only Figures 5(a) and 5(c), one would conclude that, seen in the 3 space of Channels 2, 3 and 4 the three-dimensional shape of the data structure is a flattened triangular shape. One then can conclude that the data structure forms a flattened triangular shape in 4 dimensions, and that is correct.

If one assumes that CH 1 is highly correlated to CH 2 (as it seems to be, based on Figure 5(b)) and that CH 4 is highly correlated to CH 3 (as it seems to be, based on Figure 5(c)), then the last 3 combinations of channels offer no particular surprises; they are in a manner of speaking first and second cousins of Figure 5(a). (The fact of the high correlation

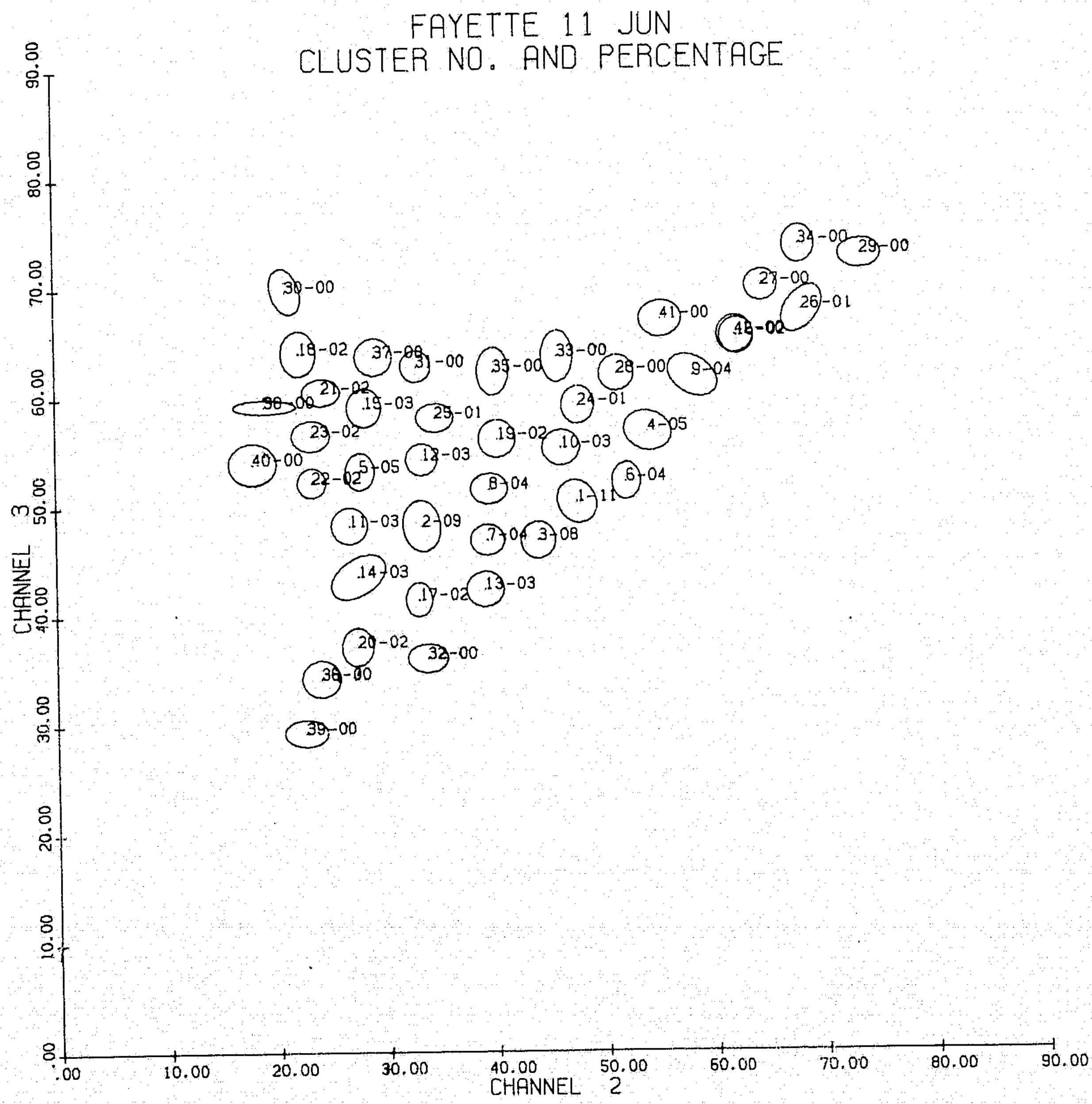


FIGURE 5(a). FAYETTE COUNTY, 1973 DATA, CLUSTER PLOT OF CHANNEL 2 VS 3

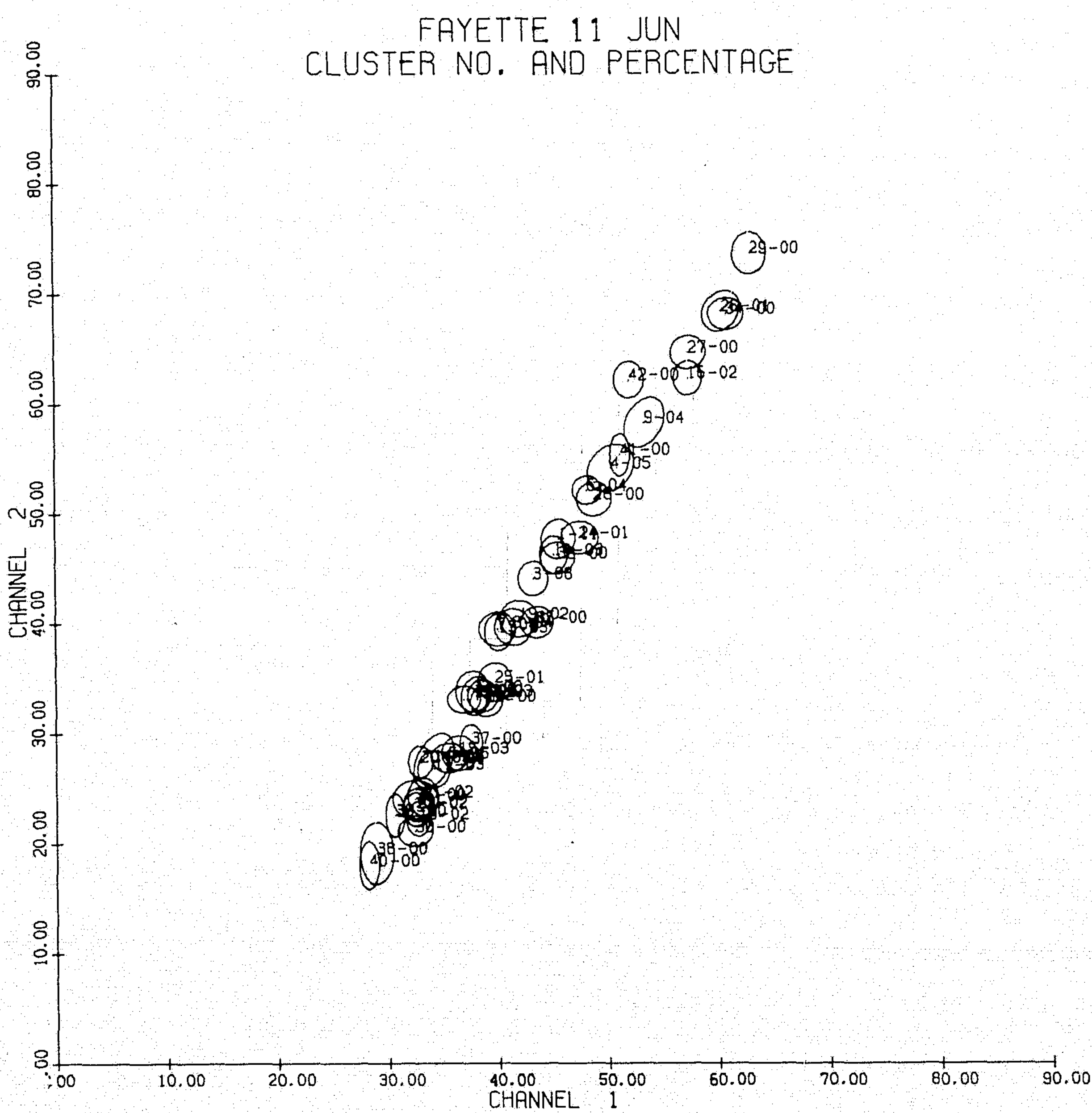


FIGURE 5(b). FAYETTE COUNTY, CLUSTER PLOT OF CHANNEL 1 VS 2

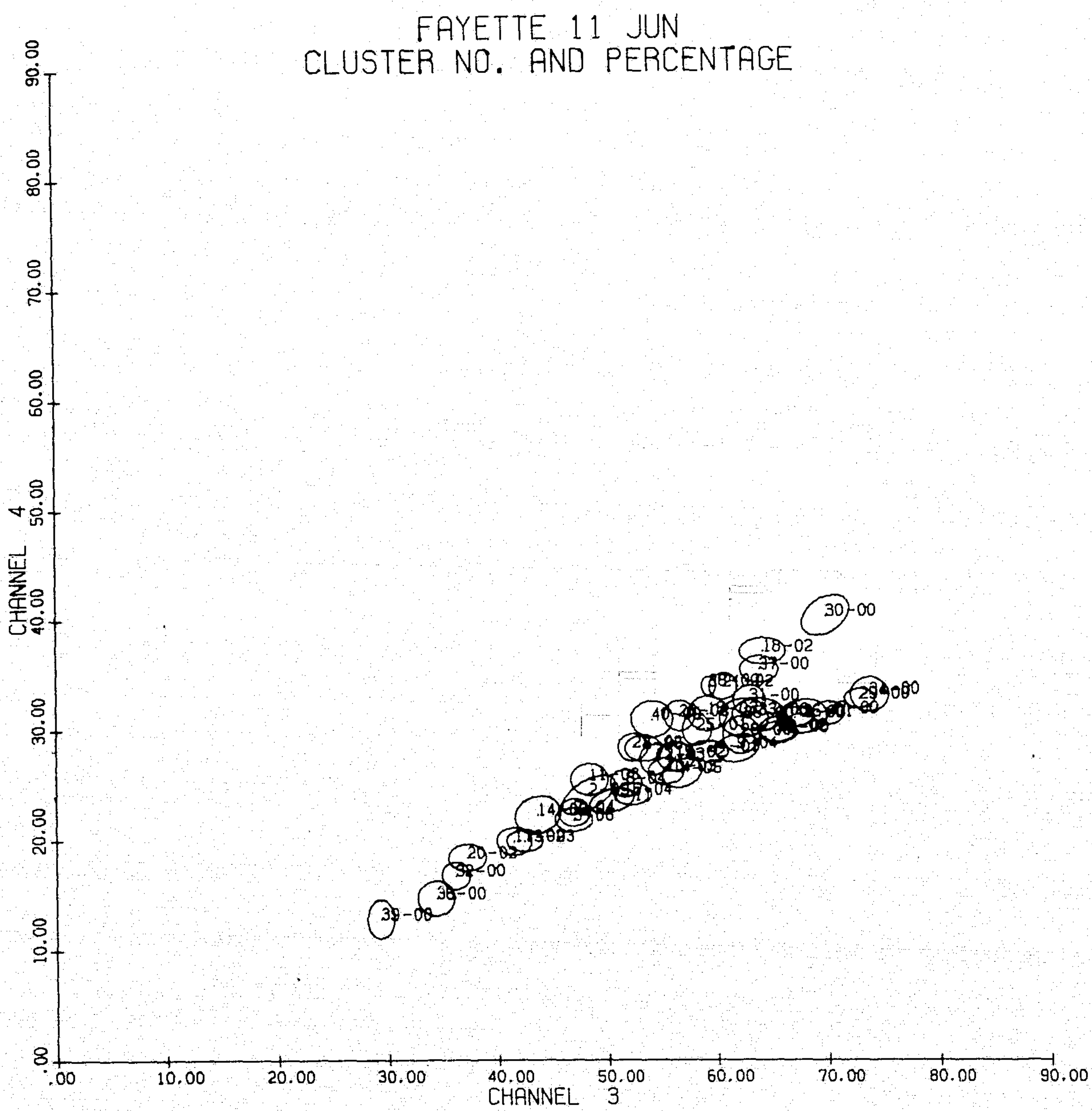


FIGURE 5(c). FAYETTE COUNTY, CLUSTER PLOT OF CHANNEL 3 VS 4

of CH 1 with CH 2 and of CH 4 with CH 3 has sometimes stimulated the comment that Landsat MSS is essentially a two-channel system; that no information would be lost by throwing away CH 1 and CH 4. On the contrary, there is significant information of several types contained in the 4 channels, as we shall see as this discussion develops.)

What is the physical reason for the data to lie in this flattened triangular structure? Figure 6 shows a model calculation of the reflectance of a crop canopy at two wavelengths, 0.65 nm and 0.75nm, corresponding to the centers of CH 2 and CH 3. The calculations were made for two soil samples, one dark, the other light, through the life of the crop. Notably, the triangular shape is outlined by the two-crop life development lines. After the crop canopy covers the soil completely, the two canopies look identical. Figure 6 is extracted from Henderson, Thomas and Nalepka, Reference [2]. The canopy model used was developed by G. Suits [3]. Roughly what seems to be occurring is that the crop starts its growth on the line of soils. As it grows, the composite reflectance of soil and crop increases the CH 3 value because of the presence of cellulose in the plant. The composite reflectance of CH 2 decreases because the Chlorophyll in the plants is highly absorbing. Hence, the radiance typical of green plants is located to the left at the tip of the triangle.

Figure 6 attempts to span the range of soil conditions by the terms "light" and "dark". Is this all there is to soils as seen in Landsat data? Condit [4,5] has measured the spectral reflectance of soil samples throughout the United States, and analyzed them in terms of their principal spectral components. We have used Condit's data to calculate the soil distribution that would be seen through the Landsat MSS spectral filters, [11]. Table 4 shows the soil reflectance mean vector and principal components in Landsat data. We will summarize those results in the following discussion.

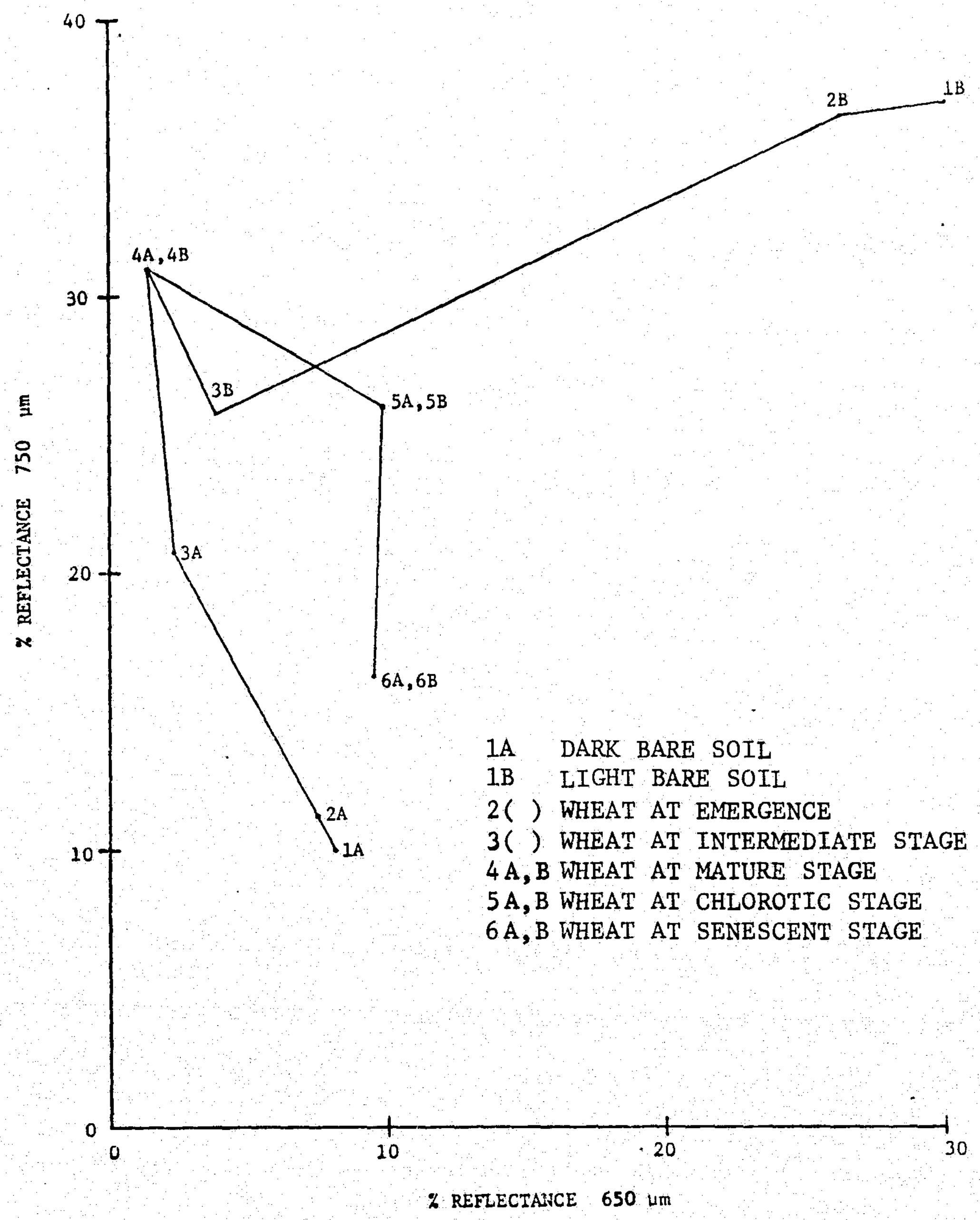


FIGURE 6. PHENOLOGY FOR WHEAT (IONIA VARIETY) BASED ON CANOPY MODEL [2]

TABLE 4. SUMMARY OF VARIOUS VECTORS AS SEEN BY LANDSAT PERCENT EFFECTIVE REFLECTANCE

	CH1	CH2	CH3	CH4	$\sqrt{\lambda}$
μ_s	15.57	21.83	25.55	31.14	48.389
V_1	14.22	17.36	18.981	20.23	35.681
V_2	4.23	.018	- 2.03	- 2.16	5.165
V_3	- .57	2.076	- 1.54	1.30	2.949
V_4	- 1.35	.166	.581	- .1408	1.486

μ_s is the mean vector of soils

V_1 through V_4 are the principal components whose amplitudes are given as $\sqrt{\lambda}$.

Figure 7 gives an idea of the distribution of soil reflectance projected into the 4 dimensions spanned by the 4 Landsat MSS Channels. That space has a "diagonal" i.e., a line along which the normalized reflectance of all channels is equal. The mean reflectance of soils lies near that diagonal. The largest principal component of soil reflectance is nearly parallel to the diagonal. The square root of the eigenvalue associated with the first component is about 35 units, (i.e., one standard deviation of the data projected onto the first principal component is about 35). The second principal component, normal to the first, has a standard deviation of about 5 units, the third of about 3 units, and the fourth of about 1-1/2 units. The unit contour ellipsoid describing the distribution of soils forms a four-dimensional flattened cigar shape, about seven times as long as it is wide, about twice as wide as it is thick, and twice as

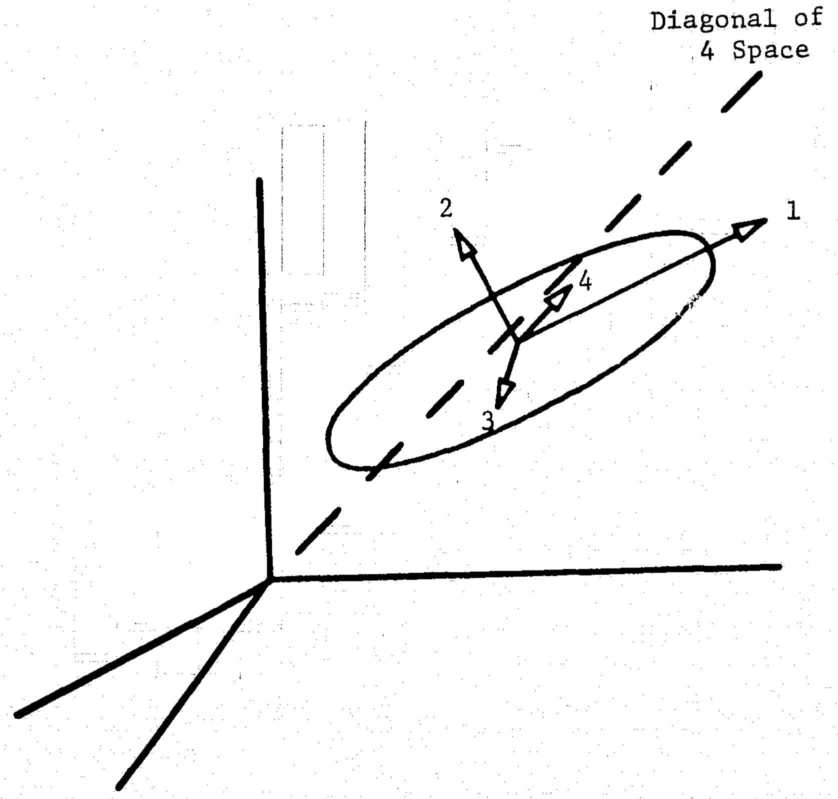


FIGURE 7. CONCEPT OF SOIL DISTRIBUTION IN 4-DIMENSIONS

thick as it is thin (which is the name for distance in the 4th direction). Hence, for some applications, we would be justified in describing the data from soil points as the "line of soils," ignoring all but the major component. In other cases, we might speak of the "plane of soils," referring to the first and second component.

Returning now to Figure 6, we notice again that after the initial development stages the two crop canopy trajectories join and fall back towards the soil line. What cannot be seen in this figure is that the line of falling back is not in the same plane (in the 4 space of Landsat data) as the two development lines up to the point where they join. The crop is yellowing, and yellow-colored things lie in a different direction away from the soil line than do green-colored things.

We now have sufficient information to create the basic image of the tasselled cap, shown in Figure 8.

The basic tasselled cap shown in Figure 8, is created by combining soil reflectance and green stuff and then adding yellow stuff. We say that the crop starts growing on the <u>plane</u> of soils. As it grows it progresses outward, roughly normal to the plane of soils, on a curving trajectory towards the region of green stuff. Next the trajectories fold over and converge on the region of yellow stuff. Finally, the crop progresses back to the soil from whence it came by any of several possible routes, depending on the crop and the harvesting practices.

Initially, we spoke of a flattened triangle. Now, we are likening the data structure to a tasselled cap. To fit both of these images the yellow point must be quite close to the side of the cap and indeed that is true. For wheat, the yellow is also accompanied by shadowing so that the yellow point is found near the dark end of the plane of soils.

The "front" of the cap looks down toward the origin of all data otherwise called THE ORIGIN. On the front of the cap is the badge of trees. Why the reflectance of trees is located just here will be explained a little further on.

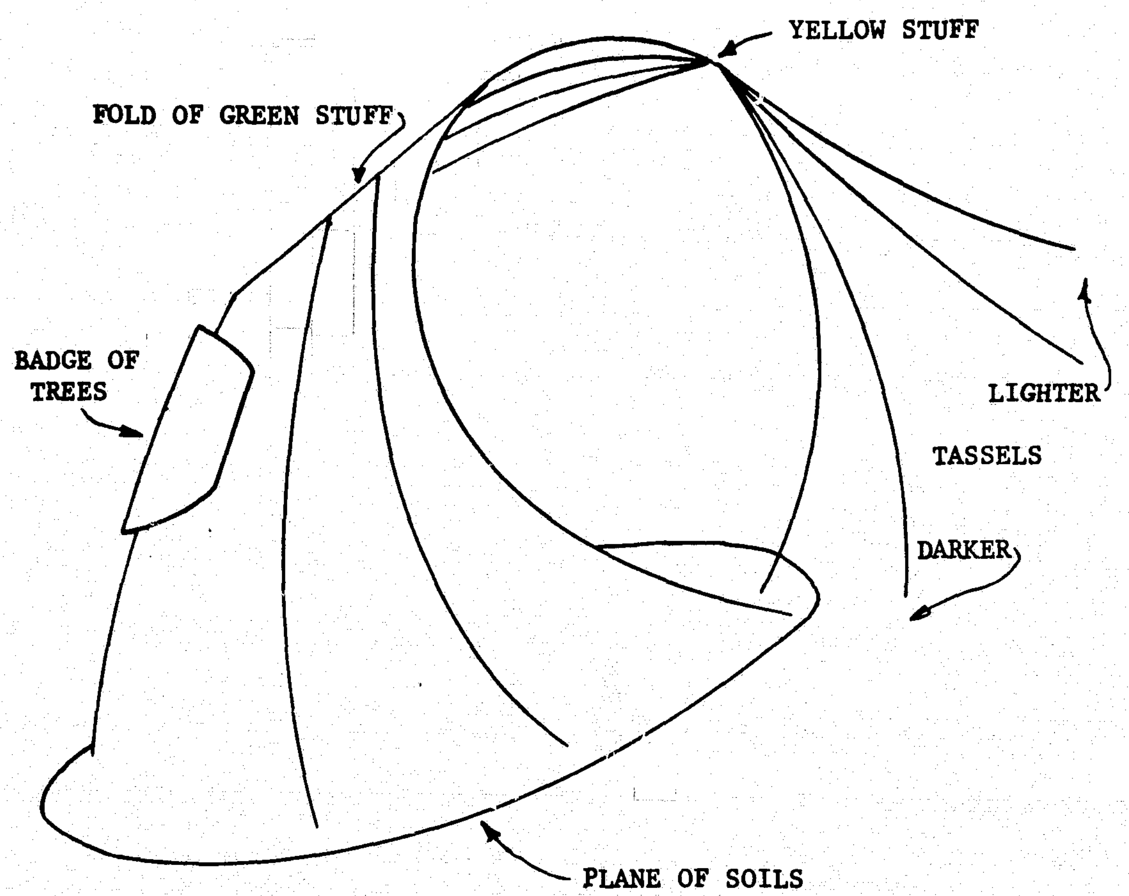


FIGURE 8. THE TASSELED CAP

Effects of Shadow

As the crop canopy develops away from the soil, the average reflectance becomes more green, but at the same time shadows develop. Initially, much of this shadow will appear on the soil portions of the composite canopy. Thus the reflectance of a crop planted on bright soil will initially migrate mainly in the direction of the origin.

A crop which is planted on dark soil will not show this behavior significantly. After all, there is little difference between the radiance of dark soils and the radiance of shadowed dark soil.

Once maximum shadowing on the soil has been reached the reflectance is more strongly influenced by the addition of green elements to the canopy. Thus the trajectory of reflectance values sweeps away from the plane of soils. Initially many of the green elements that are added are shadowed green elements. Hence, the total reflectance remains low until most of the ground is covered.

In the next stage the canopy looses most of its shadows, reaching a state of full green development. Whether a crop actually reaches this stage depends upon the planting density and upon the way its leaves form together to make a canopy.

This curving trajectory has been documented by F. Johnson [6] in Fayette County corn field data, and also has been shown in the results of a detailed modeling exercise conducted in other efforts under this contract [7,8]. Interestingly, Johnson has found that corn planted in East-West rows does not show this behavior significantly, whilst corn planted in North-South rows does show a very strong shadow effect. The reason is clear. At the time of the Landsat overpass, the Sun's rays are coming mainly from the east. Sunlight falls down the East-West rows and shadows fall on the sides of other corn plants rather than in the open rows.

Now we can see why trees occupy the place they do in reflectance space. Trees are green canopies structured so as to create a good deal of shadow.

A Fixed-Linear Transformation

It is difficult to look at Landsat data and see all of the features so far described. After all, this is a 4-dimensional space we are looking at, and it is hard to be sure we are seeing everything. Therefore, we have developed some transformations of the data which assist us to see it better.* The only one of these we will discuss at this point is a fixed affine transformation,

$$u = R^T x + r \tag{1}$$

where

- x is the LANDSAT MSS signal vector expressed in counts
- u is the transformed vector, also expressed in counts
- r is an offset vector, introduced to avoid negative values in the transformed data
- R is a unitary matrix, i.e., the columns of R are unit vectors R_1, R_2, R_3 and R_4, which are all orthogonal to each other. Superscript T indicates the transpose. Thus the application of the transformation to the data x results in a pure rotation plus a pure translation.

The components of R are chosen in the following way:

- R_1 is chosen to point along the major axis of soils in the Landsat data. A particular sample of Landsat data was chosen to derive R_1, namely Fayette County, Illinois, June 1973. Visual inspection of Figure 5(a) was used to pick out 12 soil line clusters. The best fit line to the means of those 12 clusters was chosen as the direction of R_1. R_1 is called the soil

* The transformations we have devloped have depended in part on the work of F. Johnson.[6]

brightness unit vector. The projection of a data point onto R_1 is a feature called brightness.

R_2 is chosen to point orthogonal to R_1 and toward a green cluster in the same data set. Visual inspection of Figure 5(a) was used to identify the cluster. R_2 was generated using the Gram-Schmidt orthogonalization procedure. R_2 is the green stuff unit vector. The projection of a data point onto R_2 is a feature called "green stuff."

R_3 is chosen orthogonal to both R_1 and R_2, and points toward a yellow-stuff point. There was no yellow stuff in the Fayette segment, hence an approximate spectrum of yellow corn was used to simulate or predict the yellow point in the Fayette data. Again the Gram-Schmidt procedure was used to derive the yellow-stuff unit vector.

R_4 is chosen orthogonal to R_1, R_2 and R_3. The projection of a data point onto R_4 is a feature called "non-such."

The values of R_1, R_2, R_3 and R_4 are, to the third decimal place,

$$R_1 = \begin{pmatrix} .433 \\ .632 \\ .586 \\ .264 \end{pmatrix} \qquad R_2 = \begin{pmatrix} -.290 \\ -.562 \\ .600 \\ .491 \end{pmatrix}$$

$$R_3 = \begin{pmatrix} -.829 \\ .522 \\ .039 \\ .194 \end{pmatrix} \qquad R_4 = \begin{pmatrix} .223 \\ .012 \\ -.543 \\ .810 \end{pmatrix}$$

The offset vector is arbitrary. All components equal to 32 seems to work well.

The fixed-linear transformation has several potential uses.

a. Simply by projecting the clustered data in terms of the features

of Eq. (1) we can see the data structure easily. We can also examine it to determine to what extent it actually behaves according to our imaginary picture.

b. Potentially, there is significantly less information in some of the transformed channels than in others, whereas each of the original channels is about equally information carrying. Thus, one might be able to ignore certain of the tranformed channels and this could lead to cost reduction in processing.

c. The transformation of the data allows certain diagnostic features to be extracted which are symptomatic of external effects, such as haze, H_2O vapor, illumination angle, and viewing angle.

In order to picture the data resulting from the fixed-linear transformation we show Figure 9(a) through (d), which are cluster plots of the data presented in the pairs of transformed channels. The data shown is from the Ellis County, Kansas 1TS, dated 13 June 1973. Much of the scene is bare soil. (Recall that the transformation was developed on Illinois data.) Notice that transformed channel 1 (TCH 1), which is soil brightness, and TCH 2, green stuff, contain almost all of the variation within the sample segment.

Figure 9(a) shows these two channels plotted against each other. The basic triangular shape is easily noted, now rotated to the right so that the soil line is parallel with the soil brightness axis. One noticeable effect of the transformation is to increase the apparent size of the tasselled cap, even though there was not any scale factor built in to the transformation. The reason is that in the transformed data we are seeing the tasselled cap directly from the side.

Figure 9(b) shows the yellow feature plotted versus the green feature. Notice that the data is greatly compressed in the yellow direction.

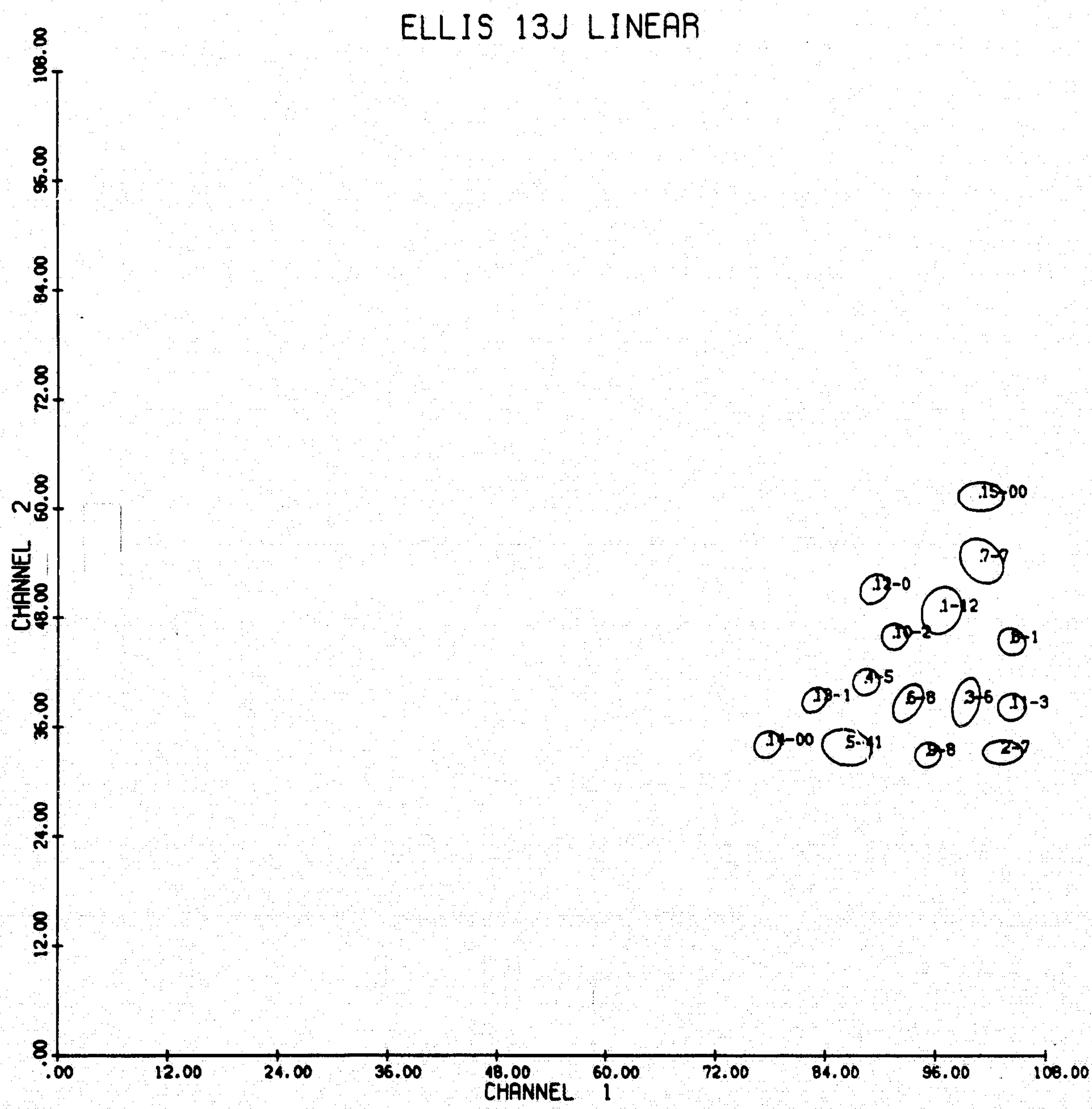


FIGURE 9(a). ELLIS COUNTY, 1974, CLUSTER PLOT OF LINEARLY TRANSFORMED DATA, CHANNEL 1 VS 2

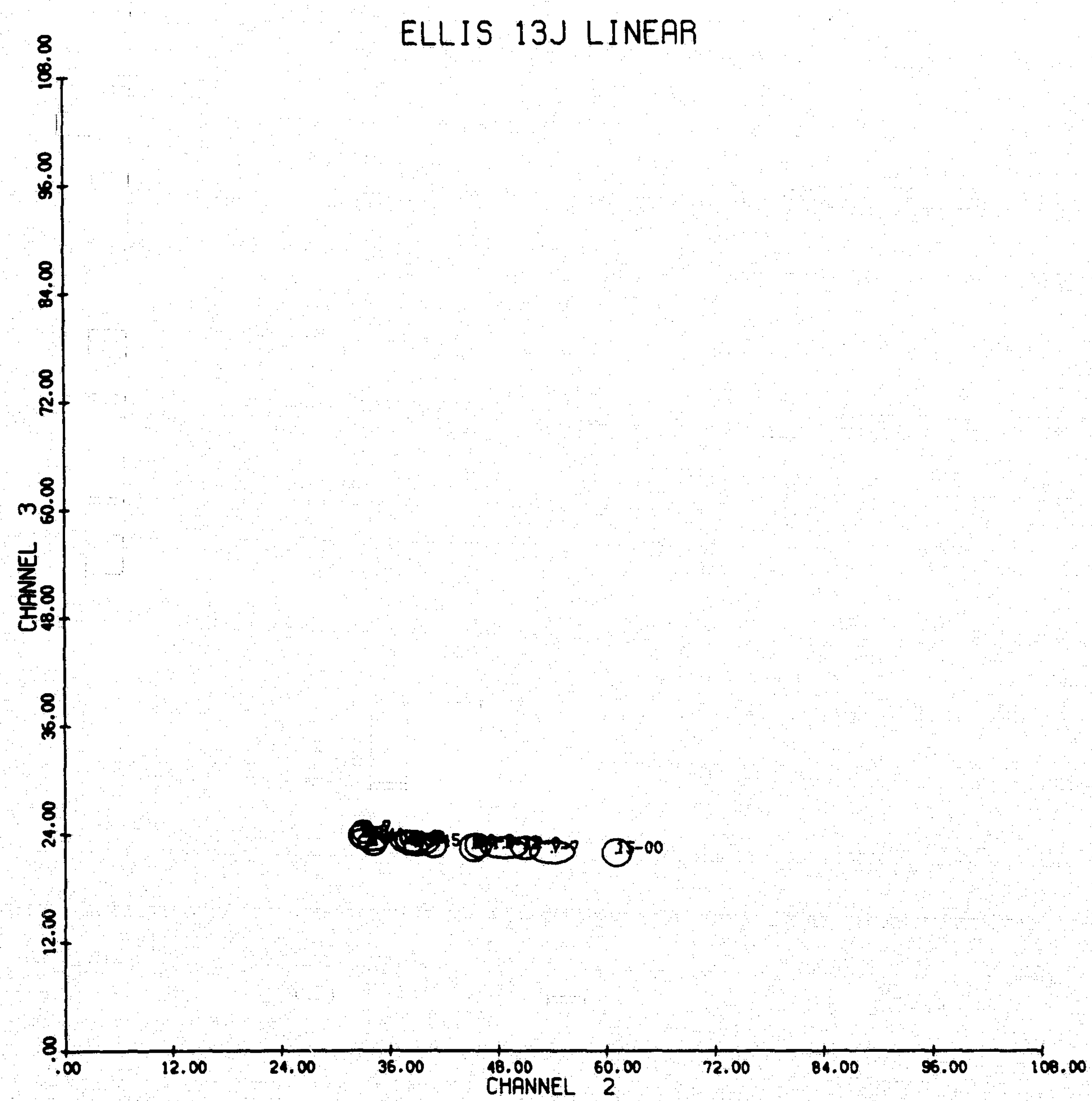


FIGURE 9(b). ELLIS COUNTY, CLUSTER PLOT OF TRANSFORMED DATA, CHANNEL 2 VS 3

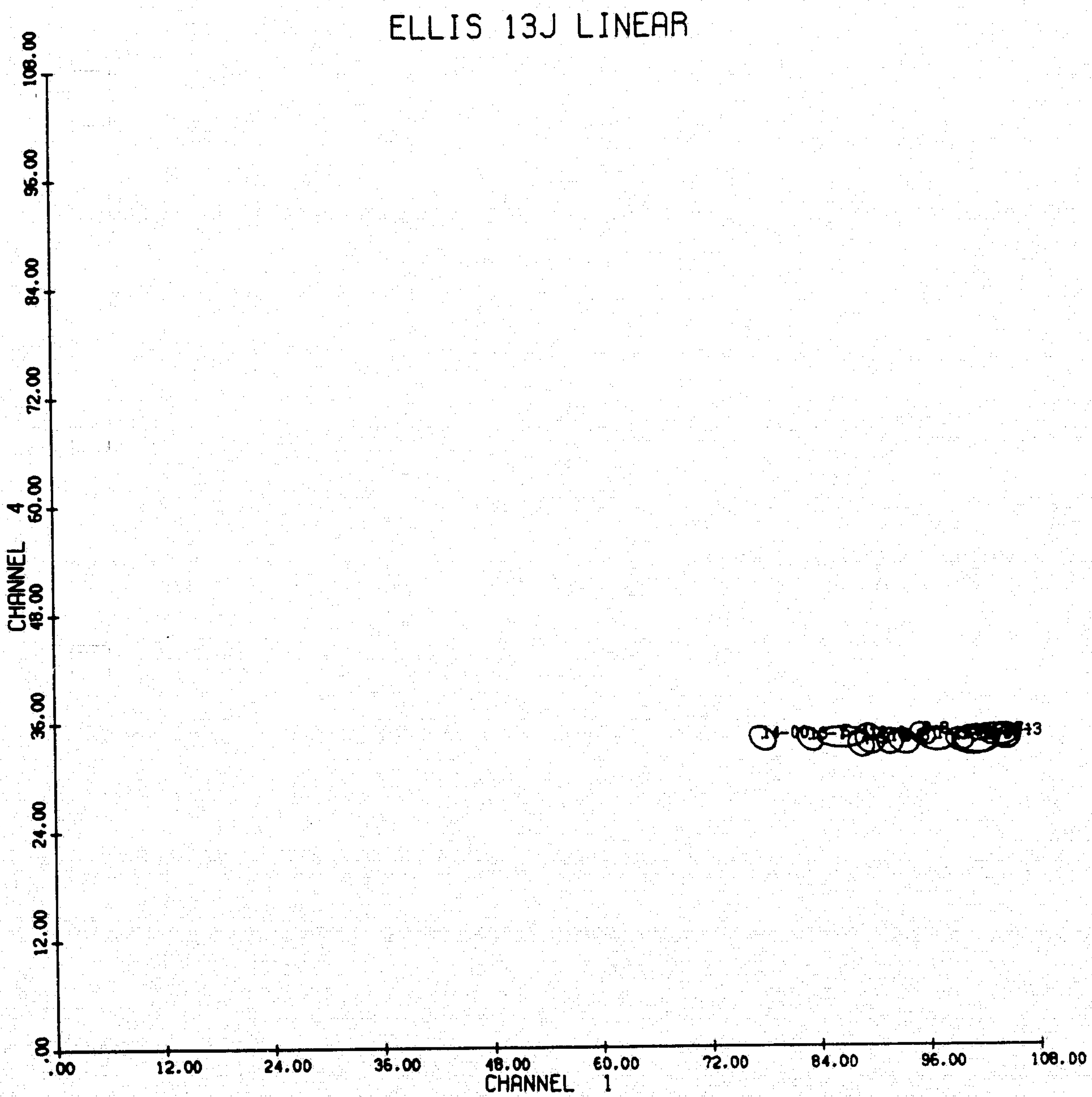


FIGURE 9(c). ELLIS COUNTY, CLUSTER PLOT OF TRANSFORMED DATA, CHANNEL 1 VS 4

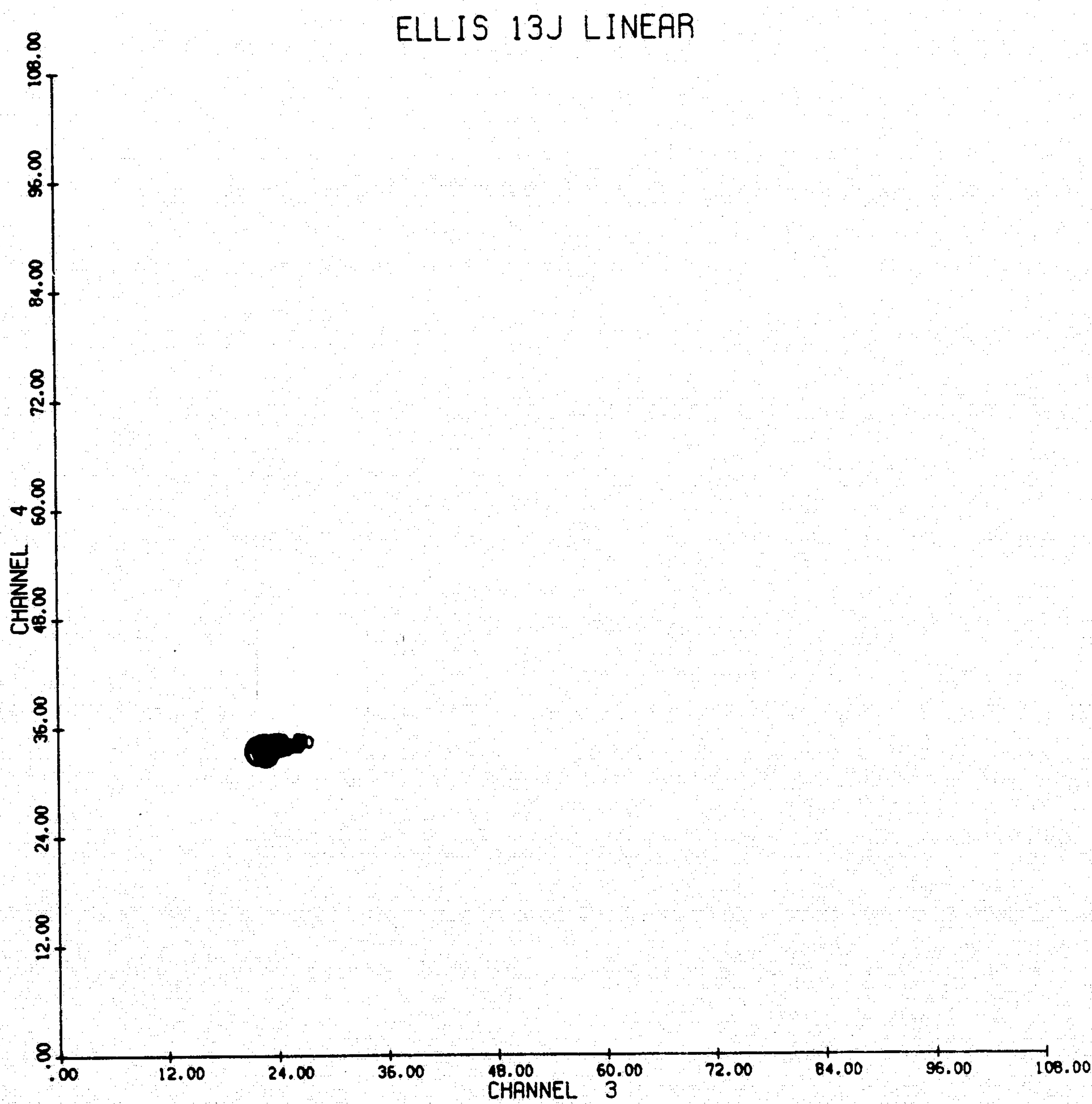


FIGURE 9(d). ELLIS COUNTY, CLUSTER PLOT OF TRANSFORMED DATA, CHANNEL 3 VS 4

Figure 9(c) shows non-such versus soil brightness. There is evidently no structure at all in the non-such direction. Figure 9(d) shows non-such versus yellow stuff. One could easily belive that these channels together carry only a tiny fraction of the information available in Landsat data. However, yellow stuff does show definite spatial structure at some times, as we will see later.

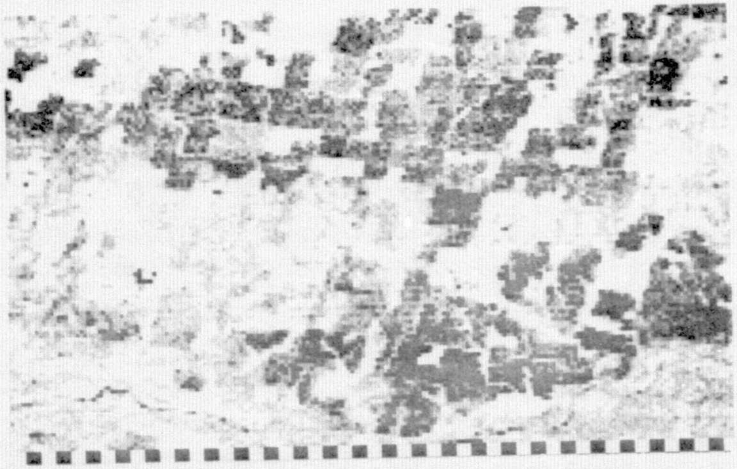
A second method of presentation of transformed data is by viewing transformed imagery. Figure 10 shows green stuff images of a LACIE sample segment in Kansas during 4 successive plant development stages.* The region at the top and bottom of the segment contains numerous winter wheat fields. The region at the center is rangeland. Figure 11 is the soil brightness image of the same data. Figure 12(b) shows non-such in the 4th biophase and is reasonably typical of non-such and yellow stuff in all of the biophases, i.e., mainly noise with almost no discernible structure. Figure 12(a) is yellow stuff in the 4th biophase. Although the dynamic range of the data is only about 10 counts, which is comparable to Figure 12(b), the strong spatial structure is evident.

Returning to Figure 10(a), we note the rangeland is somewhat green, but the fields are not green at all. The roads show up, if at all, as slightly green, due to the grass on the roadside. In Figure 10(b), the 2nd biophase, the fields show up strongly green, while the rangeland is still only somewhat green. In Figure 10(c), the 3rd biophase, both rangeland and winter wheat are green; one can imagine that the rangeland had caught up with the wheat. Finally, in Figure 10(d), the 4th biophase, the wheat is again not green. These trends are substantially what one would expect.

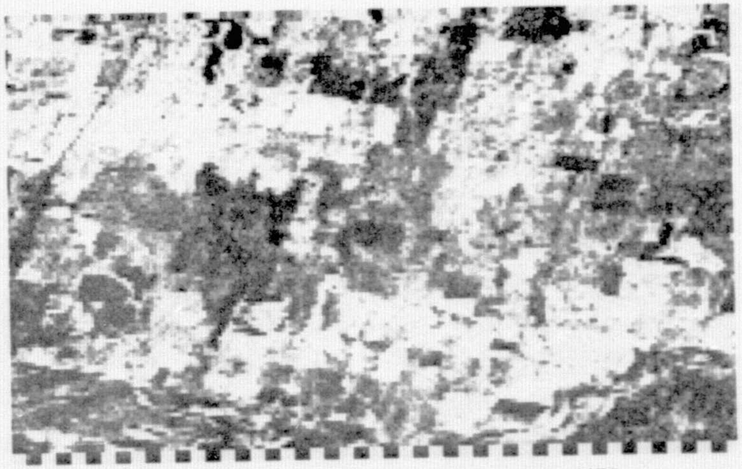
Returning to Figure 11(a), we see the soil brightness during the first biophase. One striking effect is the way the roads stand out in this image. Notice that the wheat fields are generally, but not entirely, dark.

---

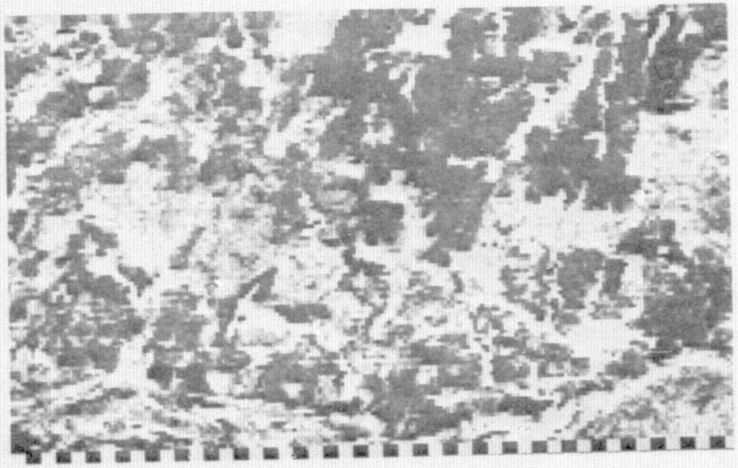
* Q. Holmes, NASA/JSC, transformed the data used in this and the next example, and created the imagery we have used in Figures 10 through 12.



(a) Biophase 1, Greenstuff, SS No. 1172



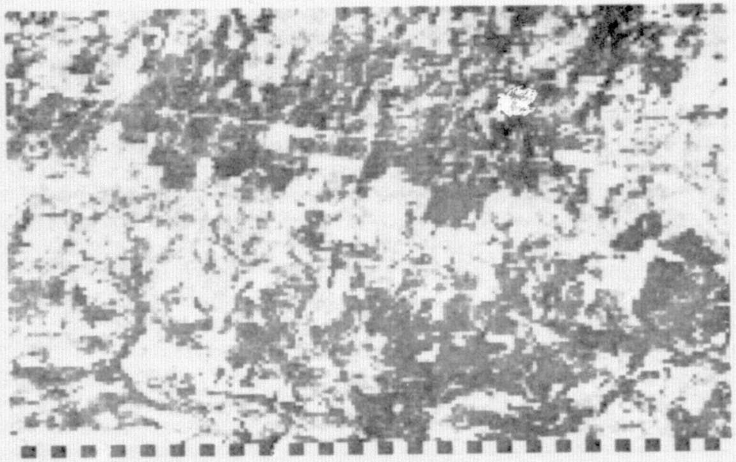
(b) Biophase 2, Greenstuff, SS No. 1172



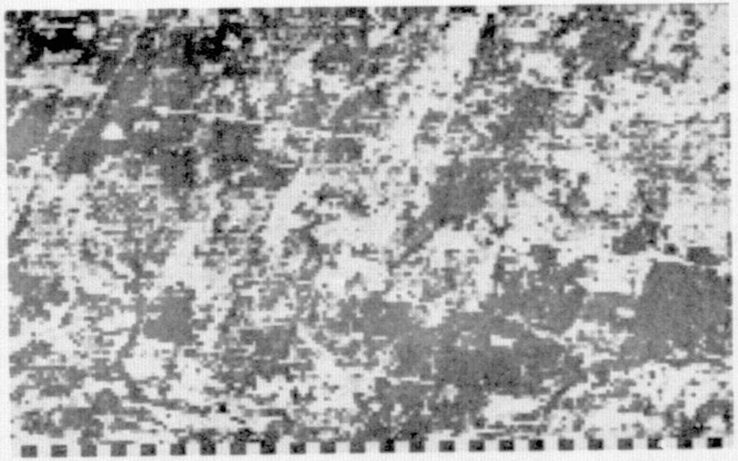
(c) Biophase 3, Greenstuff, SS No. 1172

(d) Biophase 4, Greenstuff, SS No. 1172

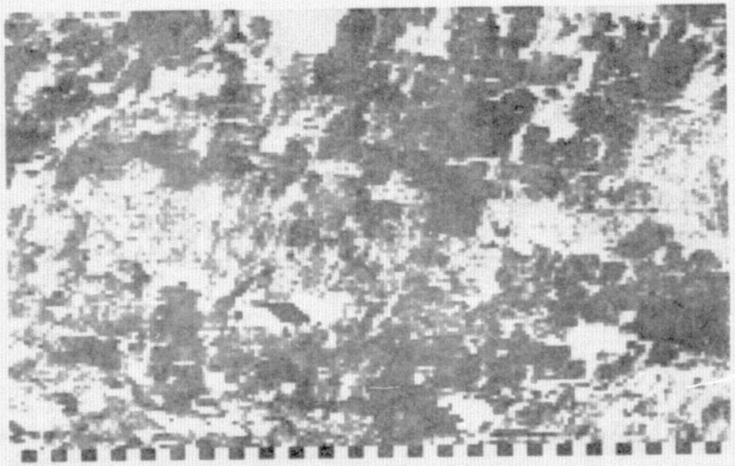
FIGURE 10. TIME PROGRESSION OF GREENSTUFF FEATURE



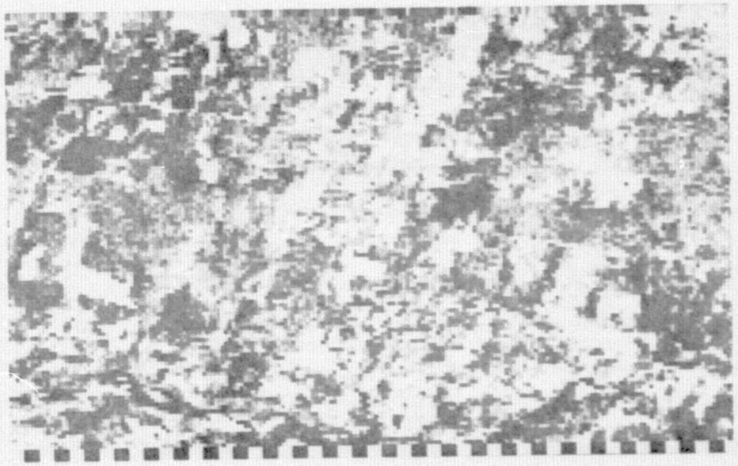
(a) Biophase 1, Brightness, SS No. 1172



(b) Biophase 2, Brightness, SS No. 1172

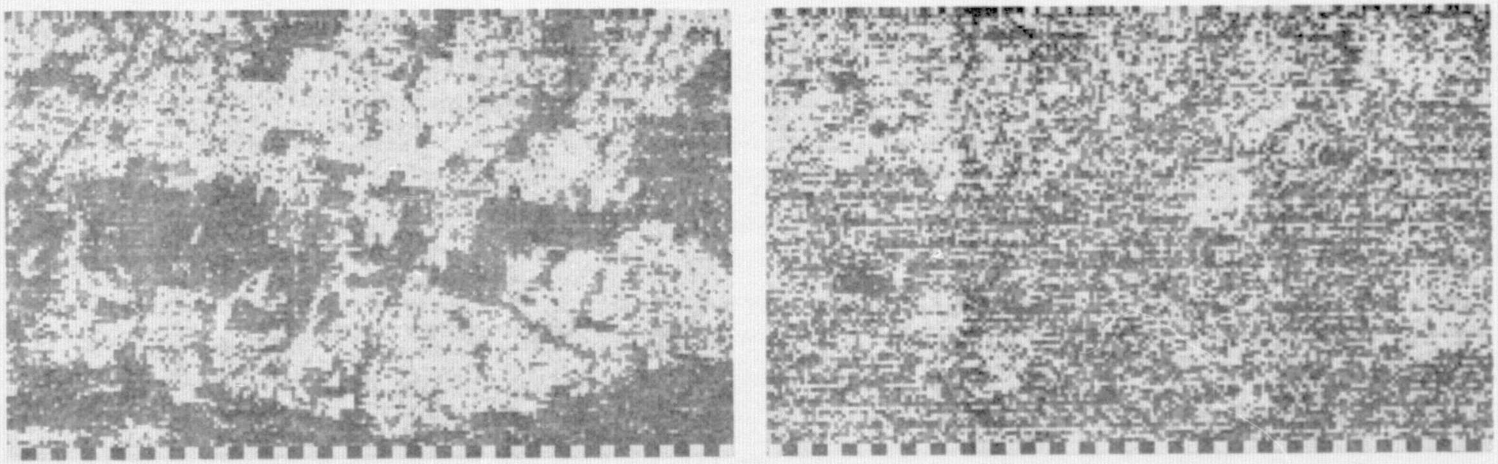


(c) Biophase 3, Brightness, SS No. 1172



(d) Biophase 4, Brightness, SS No. 1172

FIGURE 11. TIME PROGRESSION OF BRIGHTNESS FEATURE



(a) Biophase 4, Yellow Stuff, SS No. 1172

(b) Biophase 4, Non-such, SS No. 1172

FIGURE 12. SAMPLES OF YELLOW-STUFF AND NON-SUCH FEATURES

Basically these are bare soil fields and we could expect some to be light and some dark.

In Figure 11(b), the wheat fields are dark; we interpret this to mean that the fields have developed shadow in them in the process of growing. The rangeland is substantially unchanged between biophases 1 and 2, in the soil brightness feature. The roads are still bright.

In Figure 11(c), the 3rd biophase, the wheat fields have brightened up, as has the rangeland. There is little contrast between the two.

In Figure 11(d), the 4th biophase, some of the wheat fields are bright, others are not. We interpret this to mean that some are harvested (no shadows) and others are not harvested. Notice that all of the areas that appear to be wheat are yellow in the 4th biophase, but only some are bright (see Figure 12(a)). In the 4th biophase the rangeland is again moderately bright. The roads stand out by bright contrast.

Earlier, we commented that there was more than two channels' worth of information contained in Landsat data. Here we have shown an example. The green feature, the yellow feature and the brightness feature are three independent measurements. They could not all be measured and represented by a two-channel Landsat.

A third method of viewing transformed data is by looking at tables of cluster statistics or training statistics. This approach is utilized in the discussion in the following section.

The Problem of Correction for External Effects

We have discussed the Tasselled Cap as a way of integrating the spectral reflectance structure of a Landsat MSS agricultural scene. The reflectance has, for some specified conditions of observation, a corresponding radiance and a corresponding representation in Landsat counts. As the conditions of observation change, however, the relationship between reflectance and Landsat counts changes. By observation conditions we mean such items as the viewing and illumination geometry, the amount of haze in the atmosphere, the amount of H_2O vapor, the amount of cirrus

cloud and the height distribution of these in the atmosphere; also the average ground albedo in the neighborhood of the particular observed points.

Some combination of these effects is without doubt extremely significant to the problem of identifying field types in Landsat data. The very fact that the data within a local area is confined to an extremely flattened structure within the Landsat signal space makes it <u>easier</u> in a certain sense to make errors in classification. Figure 13 shows a hypothetical two-channel example in which some external effect can shift the entire set of data points sideways. Two crops, W and V, occupy a narrow region of the space, and are easily separable in that region. Assume that we train a classifier on the data from one sample segment, obtaining the signatures W and V. Then, assume that the conditions change and the entire region shifts to the position represented by W' and V'. Classification errors will now occur, but more than that, the region occupied by the original set of data points will not even include the new set of shifted data points. Figure 13 represents in an exaggerated way what really occurs due to the addition of haze to the atmosphere over a scene. The equivalent occurrence in the four-dimensional case of Landsat data would consist of a shift of the entire tasselled cap in the yellow stuff or non-such direction. Such shifts, ranging up to several standard deviations of the yellow stuff channel have been observed in randomly selected LACIE sample segments in Kansas (where standard deviation refers to the thickness of the entire tasselled cap in the yellow stuff direction).

Table 5 is a list of means and standard deviations in the yellow stuff channel and the non-such channel from several LACIE sample segments randomly selected from Kansas. These are calculated by combining clusters of both wheat and non-wheat fields. The cluster means were averaged to form, μ, a grand mean in each transformed channel and the between cluster variance component was added to the average cluster variance to obtain σ^2_{TOT}. In calculating the statistics, certain clusters were identified as

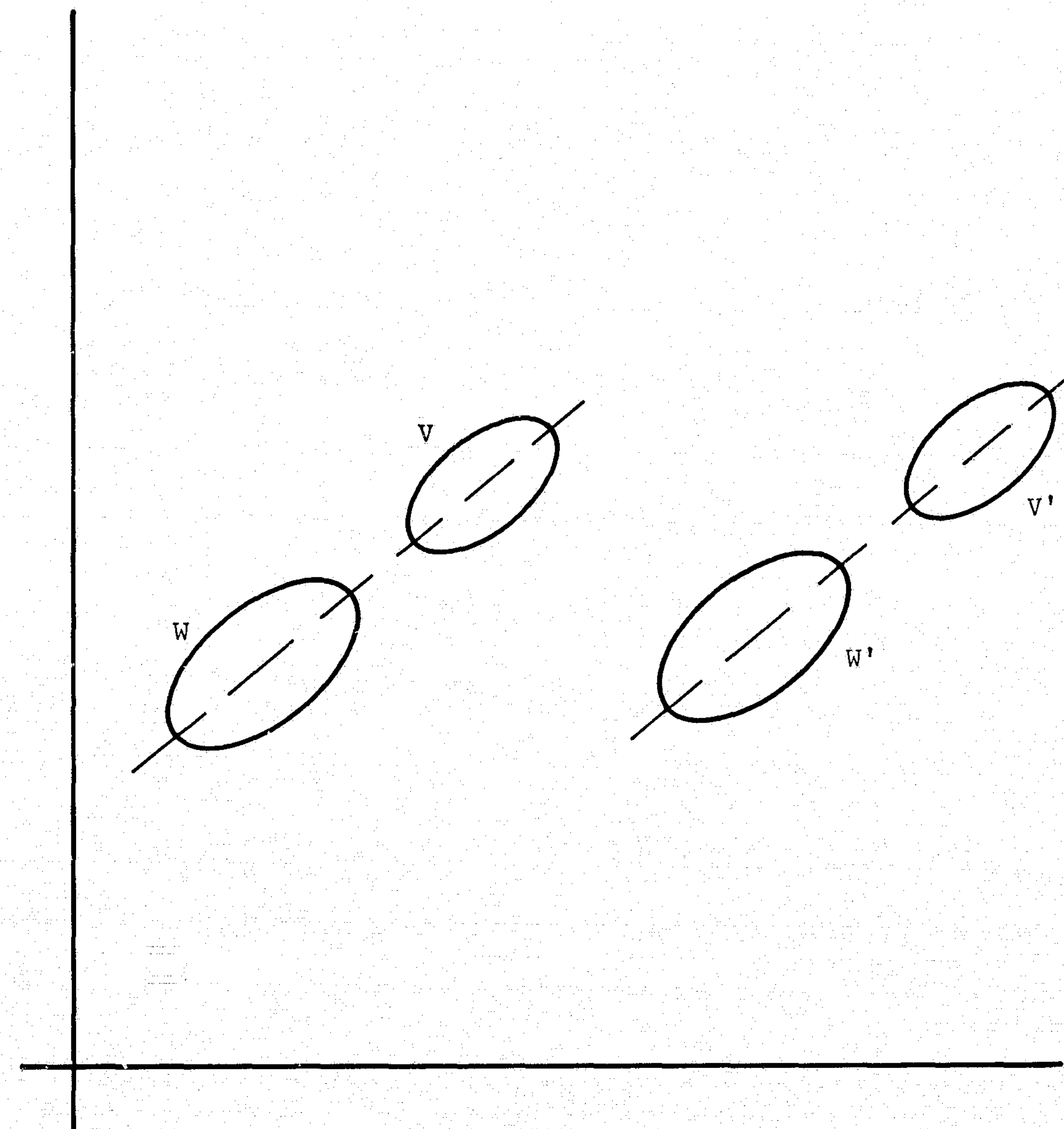


FIGURE 13. HYPOTHETICAL EXAMPLE OF HAZE EFFECT

TABLE 5. SOME EXAMPLES OF YELLOW AND NON-SUCH VARIATION FROM SITE TO SITE AND TIME TO TIME

		Yellow-Stuff		Non-Such	
Site	Time	μ	σ_{TOT}	μ	σ_{TOT}
1163	1*	22.12	2.08	33.41	1.05
	2	24.87	1.33	29.71	1.48
	3	24.04	1.61	28.46	2.08
	4	18.46	1.74	29.10	1.43
1172	1	26.12	1.21	33.23	1.06
	2	26.24	2.01	28.69	1.33
	3	24.61	1.74	28.49	1.59
	4	28.19	2.25	27.93	1.71
1854	1	24.68	1.42	33.50	1.19
	2	23.42	1.79	29.95	1.36
	3	25.84	1.38	30.59	1.48
	4	24.77	1.60	28.80	1.29
1875	1*	24.37	1.41	33.18	1.01
	2	24.51	1.51	29.07	1.62
	3*	20.39	2.39	29.25	1.44
	4	26.04	2.15	27.54	2.07
1865	1	25.84	1.41	33.10	1.72
	2	23.47	1.84	29.57	1.57
	3	26.47	2.32	27.50	1.60
	4	25.96	2.29	27.10	1.53

* Denotes that clouds were detected and the cloud clusters were omitted from the calculations.

clouds in certain passes and these were deleted from the average. The cloud clusters were detected on the basis of an algorithm developed for the PRO-CAMS effort [9] which takes advantage of a cloud's high reflectance in the bright stuff feature space and low reflectance in the yellow stuff feature space. Clouds (which may be likened to dense haze) are the darkest objects in the yellow stuff dimension. The densest clouds have a value of about zero on the yellow stuff scale. Light clouds, just discernable in imagery, have yellow stuff values in the range of 9 to 15. One particular image, sample segment 1875, pass number 3, has a gradual change from no discernible cloud at the South edge to fairly dense cloud at the North edge. The clusters deleted were all from the Northern half. Still, there is a reduced mean value (20.39) and an increased variance for that entry in the table.

We cannot discern any site-to-site or time-to-time pattern (hence, site or time dependency) in the mean values of yellow stuff. The overall standard deviation of the mean values is 2.1, whereas the average within-site standard deviation is also in the neighborhood of 2.0. Thus, there appears to be a reasonable prospect of using this feature, uncorrected, to estimate the amount or level of haze at a site. A joint use of this feature with other features for haze correction is being developed and is reported [10].

Considering non-such, the only pattern we can observe in the table is that the first pass at each site is systematically higher than the average by several counts. The sun elevation angle for these passes is about 45°, whereas for all others it is about 60°. Coincidently, all of the first passes are Landsat-1 data, while all others are Landsat-2 data. Theoretically, the non-such channel is dominated by the difference between channels 3 and 4, and ought to be influenced by the amount of water vapor present in the atmosphere, but we have no empirical evidence of that as yet.

Figure 13 also shows a shift in the brightness direction. In the real case, a negative shift in the yellow-stuff direction due to haze is also accompanied by a positive shift in brightness and a negative shift

in greenness, as well as a general contraction in scale (i.e., loss of contrast). The interactions are complicated and outside the scope of this report. The key point is that the yellow shift and the non-such seem to be diagnostic of a physical state of the atmosphere. We are attempting to exploit these diagnostic features for purposes of correcting the data for the effects of haze and viewing angle under another task of this contract, namely, signature extension. [9,10]

Imagine that a cloud represents an extreme case of haze. Then to recapitulate, a cloud would appear extremely shifted in both the negative yellow stuff and the positive brightness direction. We are continuing to experiment with a cloud detector based on this idea, i.e., if the quantity u_1-u_3 passes a certain threshold (145 works well as long as all components of the offset vector, r; are the same), a pixel is labeled cloud (u_1 and u_3 are the brightness and yellow stuff components of the vector u shown in Eq. (1).)

Point of All Shadow

AI's do not perceive the amplitude in each Landsat signal channel. Instead, they perceive preprocessed features of color and brightness. By selectively ignoring brightness they are able to concentrate on the aspect of wheat growth stage. They, thus, can ignore such confusing items as soil brightness and residual shadowing effects of view angle and illumination angle. It would seem wise to carry out the same kind of transformation on Landsat data, and express the Landsat signal in terms of "brightness", "green", and "yellow" developments, and "non-such" for purposes of developing a computer signature for wheat. (This is a nonlinear correction as opposed to the linear transformation described above.)

In principle, there exists a point, located somewhere back of the origin, called the point of all shadow. As we change viewing angle and illumination angle, the amount of shadow which can be seen in the canopy varies. The reflectance of the canopy therefore changes, becoming lighter

or darker. However, the changes are not merely towards or away from the origin, as they would be if only a change in illumination level were involved; there is also a color shift, since the radiation reflected from within the shadow region is more strongly colored than that reflected from the unshadowed region.

By making a shift of coordinates to the point of all shadow, the data can be treated as though the changes in illumination and viewing angles did not induce any color shift, but only a brightness change.

The key idea about the point of all shadow is that all points lying on any radius from this point are at the same stage of crop development. This is no doubt not perfectly true — it is in fact only an idea. But a slightly simpler version of this same idea forms the basis for the red to infrared ratio (i.e., CH 3 divided by CH 2) as a measure of green biomass.[8]

To be specific, we propose to use a transformation of the form

$$v = Q^T \frac{(x-\beta)}{s} + r \tag{2}$$

where

- x is the Landsat signal vector after haze correction
- β is the point of all shadow
- s is a brightness feature measured in the direction of soil brightness variation,

$$\left(\text{e.g.,} \quad \sum_{i=1}^{4} R_1^T (x_i - \beta_i) = s \right)$$

- Q^T is a dimension-reducing matrix such that

$$v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} \text{green-color feature} \\ \text{yellow-color feature} \end{pmatrix}$$

Thus, three features would be retained for processing, -- s and the two components of v.

In order to use this idea we have to pick a point of all shadow to work with. Several comments are in order.

a. The point of all shadow should be chosen on the extended line of soils, even if that is not truly on the reflectance diagonal. In this way, natural soil brightness variations will be lumped together with shadow variation.

b. We can pick a working shadow point and use it. If we have some success, then systematic efforts should be made to establish its position more accuractely.

c. The point of all shadow will be modified by atmosphere (haze) effects in the same way that any other point in the reflectance space will be. Any transformed features which utilize the point of all shadow as an origin will be dependent upon the haze level. Therefore, it will be necessary to carry out a correction for haze in order to properly exploit the color feature representation.

4.2.2 THE ESTABLISHMENT OF SIGNATURES

In order for the computer to make the step of suggesting field identifications to the AI, the computer must have a spectral-temporal model for field types available to it. A number of methods are available for creating such a model.

a. Baseline LACIE Signature Extension Approach

In this approach signatures from one or a few sample segments are applied to the segment being worked, after an appropriate correction for external effects such as haze or viewing angle has been made.

The main difficulty with this approach is that the sampling variance of sample segments is large. This is demonstrated in Figure 14 which shows a collection of cluster plots from sample segments in Kansas. (In these figures Landsat MSS CH2 and CH3

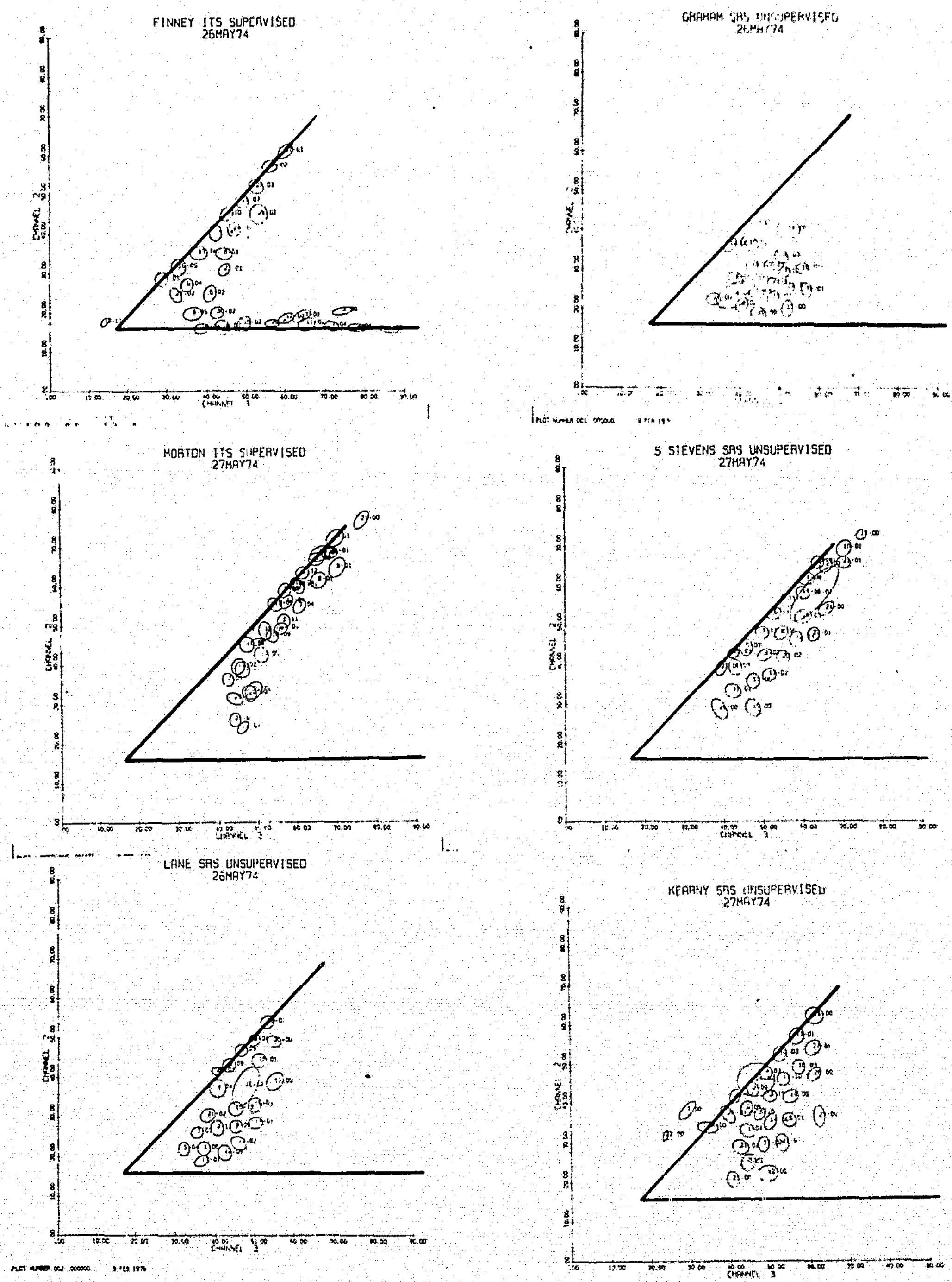
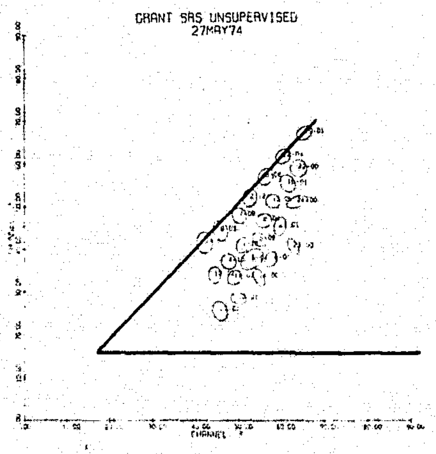


FIGURE 14. COMPARISON OF CLUSTER PLOTS (CHANNEL 2 VS 3) FOR 8 KANSAS SITES. (The solid lines are based on the Finney Site)



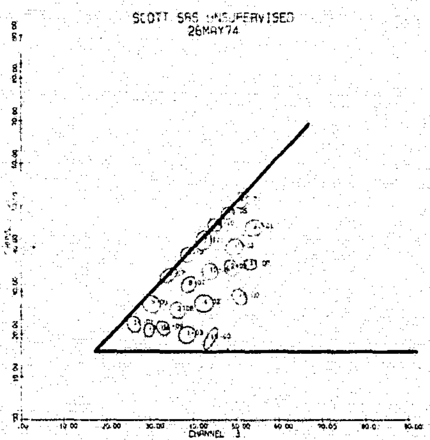


FIGURE 14. (Cont.) COMPARISON OF CLUSTER PLOTS (CHANNEL 2 VS 3) FOR 8 KANSAS SITES. (The solid lines are based on the Finney Site)

are shown reversed from Figure 4, so that the green development region is on the lower right of the graphs).

All of these passes were nominally the same biophase but are not necessarily in the same partition with respect to details of available moisture, temperature history, etc.

Attempting to extend signatures from any one sample segment to any other is equivalent to training AI's on one sample segment only. However, the training from one sample segment which contains a wide variety of crop types might be used for a "starter" model.

b. A second possible approach to building up a computer signature for wheat is to use an analytical model for the wheat spectral signature as a function of crop calendar. The model results should be checked against ground measurements and available Landsat measurements. A simulation of wheat signatures has been exercised and is reported under another task on this contract.[7]

c. A third approach, suggested by M. Trichel of NASA/JSC, involves the 3- and 4-pass combinations from the 1975 LACIE data base. Each pass would be labeled with one of 16 mini-biowindows, based on the crop calendar for the site, and each set of four passes over a site would be regarded as a 16 component (i.e., 4 passes x 4 channels) partial sample of a 64 component (i.e., 16 passes x 4 channels) vector. Data from the labeled wheat fields (from AI interpretations) would be used to estimate a 16-pass wheat signature, following the research results of T. Boullion.[16]

d. A fourth approach would be to use the delta classifier [12] to classify the blob means. (If this were employed, the haze correction and the nonlinear correction shown in Figure 3 would not be used.)

It is important to note that the system configuration shown in Figure 1(e) can be a viable one even when the quality of the computer's identification is rather poor. The AI has opportunity to disagree with the computer designation. The computer has opportunity to learn from these disagreements. Once the system is set up and operating in a local proportion estimation mode, and the AI resources required are being reduced by the computer assistance in organizing the data in the sample segment, then formal provision can be made for the learning process to occur again using results of on-going research in the SR&T community. As the computer becomes more expert at identifying blobs and clusters, the AI resources required are further reduced.

Whatever the formal learning process might be, it will be required to provide the computer with the same information that the AI has access to.

5
CONCLUSIONS AND RECOMMENDATIONS

The net result of this work to date has been the creation of a conceptual man-machine system framework for a large scale agricultural remote sensing system and the generation of some specific elements of that system. The system is based on and can grow out of the local recognition mode of LACIE, through a gradual transition wherein computer support functions supplment and replace AI functions.

Local proportion estimation functions are broken into two broad classes: organization of the data within the sample segment using spectral and spatial information within the sample segment; and identification of the fields or groups of fields in the sample segment. A set of computer programs have been implemented which assist the AI in organizing the sample segment, accept the AI's identification of fields, and tally the resulting classifications to produce a proportion estimate. The central computer programs which accomplish these results are modifications of existing ERIM programs. Thus, spectral-spatial clustering is accomplished by modifying an existing clustering algorithm to accept line and point numbers as channels. Blob clustering is accomplished by modifying the same program to accept and cluster the means of blob rather than pixel values. Another existing program was used to strip the boundary pixels from blobs. A special program had to be written to organize the blob tables. An existing program was used to tally the area of wheat inside of a sample segment, given the blob or cluster identifications.

A few examples of the exercise of this function have been produced, and await critical evaluation.

The structure of Landsat data has been explored with the objective of establishing a conceptual basis for computer identification of crops. A heuristic view of the spectral-temporal structure of Landsat data is described in the Tasselled Cap. This heuristic idea is used to describe

a system incorporating several interlinked steps, — cloud removal, haze correction, nonlinear feature extraction and classification according to a signature model, — in support of local field identification by AI and by computer. Possible methods of obtaining a signature model for wheat are outlined.

With respect to the first broad function, sample segment organization and proportion estimation, there are numerous possible variations on the theme. We have implemented one of them and it is well enough defined to be implemented elsewhere as a research tool. The most immediate need is to allow working AI's to obtain some experience with this tool. This might be accomplished by modifying existing utility programs at Johnson Space Center and writing others to accomplish these functions. To be implemented for formal test and evaluation for possible inclusion in LACIE, the approach would have to be specified at a greater level of detail than given in this report. However, the specfication would not require an inordinate additional effort.

With respect to the second broad function, field identification, many degrees of implementation are possible. All demand some measure of correction for external effects and this is in the process of being implemented. The implementation of a starter signature model which can be used to tentatively identify blobs is a short next step. However, the institution of an AI-computer interactive loop in which the signature is gradually improved due to the recorded corrections from the AI is a longer lead time function. We recommend that research on this topic be undertaken in the SR&T program soon. The analysis should include ancillary data available both to AI and to the computer, especially parameters relating to crop calendar, as part of the definition of signature.

REFERENCES

1. Gupta, J.N. and P.A. Wintz, Multi-Image Modeling, TR-EE-74-24, School of Electrical Engineering, Purdue University, West Lafayette, Indiana, September 1974.

2. Henderson, R.G., G.S. Thomas and R.F. Nalepka, Methods of Extending Signatures and Training without Ground Information, Appendix 3, 109600-16-F, Environmental Research Institute of Michigan, Ann Arbor, Michigan, May 1975.

3. Suits, G.H., The Calculation of the Directional Reflectance of a Vegetative Canopy, Proceedings of the Eighth International Symposium on Remote Sensing of Environment, Volume II, Ann Arbor, Michigan, 1972, pp. 117-125.

4. Condit, H.R., The Spectral Reflectance of American Soils, Photogrammetric Engineering, Vol. 36, September 1970, p. 955.

5. Condit, H.R., private communication.

6. Johnson, F., CITARS, Volume VIII, Data Processing at the Earth Observations Division, Lyndon B. Johnson Space Center, Part 5, Fayette County, Illinois Supplement: Graphic Study of Corn and Soybean Data, JSC-09391, December 1975.

7. Malila, W.A., R.C. Cicone, and J.M. Gleason, Wheat Signature Modeling and Analysis for Improved Training Statistics, 109600-66-F, Environmental Research Institute of Michigan, Ann Arbor, Michigan, May 1976.

8. Colwell, J.E. and G.H. Suits, Yield Prediction by Analysis of Multispectral Scanner Data, 109600-17-F, Environmental Research Institute of Michigan, Ann Arbor, Michigan, May 1975.

9. Erickson, J.D. and R.F. Nalepka, PROCAMS: A Second-Generation Multispectral-Multitemporal Data Processing System for Agricultural Mensuration, Proceedings of the Third Symposium on Machine Processing of Remotely Sensed Data, West Lafayette, Indiana, June 1976.

10. Lambeck, P.F. and D.P. Rice, Signature Extension Using Transformed Cluster Statistics and Related Techniques, 109600-70-F. Environmental Research Institute of Michigan, Ann Arbor, Michigan, May 1976.

REFERENCES (CONT.)

11. Kauth, R., The Tasselled Cap Revisited, JSC Memorandum TF3-75-5-190, 13 May 1975.

12. Engval, J. and J. Tubbs, A Non-Parametric Approach to Classifying Remotely Sensed Data, Proceedings of the Symposium on Machine Processing of Remotely Sensed Data, West Lafayette, Indiana, June 1976.

13. Bellman, R., Introduction to Matrix Analysis, McGraw Hill, New York, 1960, p. 44.

14. Potter, J. and M. Shelton, Effect of Atmospheric Haze and Sun Angle on Automatic Classification of ERTS-1 Data, Proceedings of the Ninth International Symposium on Remote Sensing, April 1974.

15. Classification and Mensuration Subsystem (CAMS) Requirements, LACIE-00200, Vol. II, National Aeronautics and Space Administration, Lyndon B. Johnson Space Center, Houston, Texas, December 1974

16. Boullion, T.L., Some Results on Randomly Missing Data In Discriminant Analysis, Annual Report, University of Texas at Dallas, JSC-09703, May 1975, pp. 94-104.

APPENDIX I

CROP CODES

Code	Crop
100	Spring Wheat (General)
101 → 126	Spring Wheat Varieties
200	Barley (General)
201 → 216	Barley Varieties
300	Oats (General)
301 → 307	Oat Varieties
400	Winter Wheat (General)
401 → 424	Winter Wheat Varieties
500	Grasses/Pasture
600	Other Crops (General)
601 → 618	Other Crops (Specific)
700	Summer Fallow
800	Non-Agricultural

Field No.	ID	Field No.	ID	Field No.	ID
1	700	13	400	26	700
2	700	14	700	27	700
3	404	15	402	28	500
4	402	16	400	29	400
5	700	17	400	30	700
7	402	18	500	31	400
8	700	19	700	32	402
9	402	20	616	33	700
10	700	21	616	34	404
11	700	23	700	35	700
12	400	25	404	36	616

Field No.	ID	Field No.	ID	Field No.	ID
37	408	82	607	123	500
39	404	83	616	124	500
40	400	85	700	125	500
41	402	86	700	126	402
42	700	87	402	127	700
43	400	88	607	128	700
44	400	89	408	129	402
46	400	91	408	130	607
47	700	93	404	131	402
48	616	94	404	134	408
49	408	97	606	135	607
51	500	98	700	137	700
52	400	99	408	139	402
53	500	102	700	140	700
54	400	103	616	141	700
55	407	106	616 *	144	400
56	404	106	607 *	147	402
58	700	108	402	148	500
59	500	110	402	149	500
60	616	111	607	150	409
61	607	113	409	151	607
62	616	114	700	152	408
63	402	115	402	153	402
64	700	116	700	154	607
66	402	117	402	155	408
73	700	118	700	158	616
74	700	119	607	159	700
75	700	120	616	160	700
76	700	121	607	161	700
77	404	122	700	163	700

* This is the way it is shown in our ground information.

Field No.	ID	Field No.	ID
164	500	209	616
166	700	210	700
167	400	211	600
168	607	214	602
169	607	215	602
170	607	217	500
171	402		
173	700		
174	607		
175	700		
176	700		
177	400		
179	607		
180	607		
182	700		
185	400		
186	400		
187	400		
188	607		
189	607		
192	607		
193	607		
194	700		
196	402		
198	402		
200	400		
201	607		
206	607		
207	700		
208	616		